

**Transcriptome analysis of *Hamelia patens* (Rubiaceae)  
anthers reveals candidate genes for tapetum and pollen wall  
development**

Lin YUE<sup>1,2</sup>, David TWELL<sup>3</sup>, Yanfeng KUANG<sup>1,\*</sup>, Jingping LIAO<sup>1</sup> & Xianqiang ZHOU<sup>4</sup>

<sup>1</sup> Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing, China

<sup>3</sup> Department of Genetics, University of Leicester, Leicester LE1 7RH, United Kingdom

<sup>4</sup> Beijing Genomic Institute (BGI) at Shenzhen, Shenzhen, China

**\*Corresponding author:** Yanfeng KUANG (E-mail: [kuangyf05@scbg.ac.cn](mailto:kuangyf05@scbg.ac.cn))

## Abstract:

Studies of the anther transcriptome on non-model plants without a known genome are surprisingly scarce. RNA-Seq and digital gene expression (DGE) profiling provides a comprehensive approach to identify candidate genes contributing to developmental processes in non-model species. Here we built a transcriptome library of developing anthers of *Hamelia patens* and analysed DGE profiles from each stage to identify genes that regulate tapetum and pollen development. In total 7,720 putative differentially expressed genes across four anther stages were identified. The number of putative stage-specific genes was: 776 at microspore mother cell stage (MMC), 807 at tetrad stage (TET), 322 at uninucleate microspore stage (UNM) and the highest number (1,864) at bicellular pollen stage (BCP). GO enrichment analysis revealed 243 differentially expressed and 108 stage-specific genes that are potentially related to tapetum development, sporopollenin synthesis and pollen wall. The number of expressed genes, their function and expression profiles were all significantly correlated with anther developmental processes. Overall comparisons of anther and pollen transcriptomes with those of rice and Arabidopsis together with the expression profiles of homologs of known anther-expressed genes, revealed conserved patterns and also divergence. The divergence may reflect taxon-specific differences in gene expression, the use RNA-seq as a more sensitive methodology, variation in tissue composition and sampling strategies. Given the lack of genomic sequence, this study succeeded in assigning putative identity to a significant proportion of anther-expressed genes and genes relevant to tapetum and pollen development in *H. patens*. The anther transcriptome revealed a molecular distinction between developmental stages, serving as a resource to unravel the functions of genes involved in anther development in *H. patens* and informing the analysis of other members of the Rubiaceae.

**Keywords:** Transcriptome, Rubiaceae, *Hamelia patens*, Anther, Tapetum, Pollen wall

## Introduction

Pollen grains are the microgametophytes of seed plants that produce the male gametes required for sexual reproduction (Borg and Twell, 2011). Although the overall structure of pollen is conserved in angiosperms, pollen shows considerable variation in size, shape and surface characteristics (Heslop-Harrison, 1968; Blackmore et al., 2007). Recently, the diversity and evolution of palynological characters were documented across the angiosperms based on the most comprehensive classification APG III (2009). Many characters, such as exine structure and aperture features, have indicated key evolutionary transitions in pollen morphology and have proven to be informative at different taxonomic levels (Wortley et al., 2015; Lu et al., 2015). However, little is known about the genetic basis underlying pollen evolution. Comparative analysis has indicated significant conservation in anther gene expression patterns between *Arabidopsis* and maize (Ma et al., 2008). Further comparative analysis of anther and pollen transcriptome profiles for diverse angiosperms would complement understanding the conservation of the molecular mechanisms underlying pollen development (Rutley and Twell, 2015; Wilson and Zhang, 2009).

Pollen development involves the coordination of cellular activities and the underlying gene expression in sporophytic and gametophytic cells of the anther (Ariizumi and Toriyama, 2011). Anthers are composed of developing gametophytes and a number of surrounding cell layers: the innermost tapetum, the middle layer, the endothecium and the outer epidermis. Initially, archesporial cells form in the L2 layer of the anther primordium and divide periclinally to form outer primary parietal cells and inner primary sporogenous cells. The primary sporogenous cells undergo divisions to form microspore mother cells (MMCs), whereas the primary parietal cells (PPCs) undergo a series of divisions to form the cell layers of the anther wall (Scott et al., 2004). The MMCs undergo meiosis producing tetrads of haploid microspores at tetrad stage (TET). The tetrads are usually separated as free microspores after the callosic walls are dissolved by callase at the uninucleate microspore stage (UNM). The free microspores divide asymmetrically to segregate the male germline and develop further into pollen grains before release at bicellular pollen stage (BCP) or tricellular pollen stage (TCP) (McCormick, 1993; Twell, 2011). Pollen development relies on the complex interactions between reproductive and non-reproductive tissues of the anther, especially the tapetum, which provides enzymes for the dissolution of tetrads, nutrients to the developing microspores and materials deposited onto the pollen exine (Goldberg et al., 1993; Tsuchiya et al., 1994; Li et al., 2006).

A significant number of key genes required for anther and pollen development have been identified by forward genetic screens in model and crop plants. These genes, which include a significant number of transcriptional regulators, control formation and degeneration of the tapetum, microspore release from the tetrad and formation of the complex pollen wall (Boavida et al., 2005; Wilson and Zhang, 2009; Shi et al., 2015). Key genes involved in tapetum formation and differentiation include,

*EXCESS MICROSPOROCYTES1 / EXTRA SPOROGENOUS CELLS (EMS1/EXS)* (Zhao et al., 2002), and *TAPETUM DETERMINANT 1(TPD1)* (Yang et al., 2003). Genes involved in the programmed cell death (PCD) of tapetum cells include *TAZI* (Kapoor et al., 2002), *TAPETUM DEGENERATION RETARDATION (TDR)* (Li et al., 2006), *PERSISTENT TAPETAL CELL 1 (PTC1)*, *API5* (Li et al., 2011), *Osc6* (Zhang et al., 2010) and *CEP1* (Zhang D. D. et al., 2014). The absence of these genes results in abnormal tapetal PCD and sterile pollen.

Transcriptional regulation is a major mechanism controlling anther development in Arabidopsis. The bHLH transcription factor *DYSFUNCTIONAL TAPETUM (DYT1)* plays a critical role in regulating tapetum function and pollen development (Zhang et al., 2006; Feng et al., 2012; Zhu et al., 2015). DYT1 regulates the expression of the *ABORTED MICROSPORES (AMS)* (Sorensen et al., 2003; Xu et al., 2014), *MALE STERILITY1 (MS1)* (Vizcay-Barrena and Wilson, 2006), and other tapetum preferential genes, primarily via *DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION 1 (TDF1)* (Gu et al., 2014). Important regulators of tetrad callose degradation include the Arabidopsis transcription factor gene *AtMYB103* (Higginson et al., 2003; Zhang et al., 2007), and the rice  $\beta$ -1-3, glucanase encoded by *Osg1* (Wan et al., 2011). Key genes required for pollen exine development have also been characterized, including *MS2* (Aarts et al., 1997), *ACOS5* (Souza et al., 2009), *CYP703A2* (Morant et al., 2007), *CYP703A3* (Aya et al., 2009) and *EFD* (Hu et al., 2014). Moreover, recent co-expression analysis revealed that AMS acts as a master regulator coordinating pollen wall development and sporopollenin biosynthesis in Arabidopsis (Xu et al., 2014). Thus, current understanding of anther regulatory genes is overwhelmingly derived from studies of model plants and crops and little information about gene activities in developing anthers is known from non-model species.

Transcriptome analyses have significantly enriched knowledge of the repertoire of genes expressed during pollen development. Genome-scale analyses have been described for developing anthers and pollen, germinating pollen and pollen tubes, and mature pollen transcriptome data for at least 10 different species (Huang et al., 2011; Rutley and Twell, 2015). Transcriptome profiling at different developmental stages have been reported for Arabidopsis (Honys and Twell, 2004), rice (Wei et al., 2010), and tobacco (Bokvaj et al., 2015). Transcriptome studies of developing Arabidopsis anthers and pollen have revealed genes that function as early as meiosis, in the tapetum and in exine formation (Dukowic-Schulze and Chen, 2014). Recently, genome-wide co-expression analysis revealed 98 candidate genes closely associated with pollen wall development (Xu et al., 2014). Moreover, comparative studies in rice involving microarray studies of developing anthers have also revealed complex transcriptomes (Deveshwar et al., 2011).

RNA-Seq based methods have led to a dramatic acceleration in gene discovery (Barrero et al., 2011; Garg et al., 2011; Shi et al., 2011), which has rapidly broadened understanding of the complexity of gene networks and their regulation (Wang et al.,

2010; Hua et al., 2011). Moreover, there are limitations of microarray technology such as lack of quantitative detection of transcripts with different probe sets, low abundance transcripts are difficult to detect, and a reference genome and/or expressed sequence tags (ESTs) are required (Huang et al., 2011; Zhao et al., 2014). Studies of anther transcriptomes based on RNA-seq are limited to those in a few crops and model plants, however, expression profiling of developing anthers by Next Generation Sequencing (NGS) from species whose genomes have not been sequenced is anticipated (Wang et al., 2009; Huang et al., 2011).

Rubiaceae, commonly known as the coffee family, is the fourth largest in the angiosperms and has important economic utilities; in addition to the valuable beverage crop coffee, as a source of quinine, dyes and ornamentals. It is a eurypalynous family with great diversity in pollen size and shape, aperture type, number and position, pollen wall stratification and sexine ornamentation (Dessein et al., 2005). Some pollen features have significant value in stimulating further palynological investigations, for example, the development of Ubisch bodies provides a potential model for studying sporopollenin deposition (Dessein et al., 2005). However, there is limited information about the gene expression patterns underlying anther and/or pollen development in Rubiaceae. Current molecular data from model species and a few crop plants are insufficient to resolve the mechanisms that regulate the development, function and evolution of angiosperm pollen. More attention to such a palynologically well-documented family would help clarify the molecular processes underlying anther/pollen development in angiosperms and the associated evo-devo mechanisms.

In this paper, *Hamelia patens* Jacq. was chosen for study as a member of the coffee family. *H. patens* originates in tropical America and is a popular ornamental plant in south and southwest China. The ease of cultivation and long blooming period from May to October enables facile collection of floral material. Further, this species produces triaperturate pollen grains, representing the plesiomorphic condition in the Rubiaceae family. These advantages make *H. patens* a potentially valuable system for developmental and molecular research in Rubiaceae.

Transcriptome profiles of *Hamelia patens* developing anthers from microspore mother cell to bicellular pollen grain stages were investigated using Next Generation Sequencing (NGS) and NGS-based digital gene expression (DGE) tag profiling. The aims of the present study were, (1) to provide molecular insight into anther and pollen development of this phylogenetically and palynologically important family, and (2), to reveal novel candidate genes involved in tapetum and pollen development, to shed light on the evo-devo mechanisms of angiosperm pollen development.

## **1. Materials and Methods**

### **1.1 Sample preparation for anther development**

Flowers of *Hamelia patens* Jacq. at various developmental stages were harvested from plants growing in South China Botanical Garden (SCBG), Guangdong Province,

China. Each flower contains five anthers. The length of flowers and the anthers was measured using a stereo microscope. The classification of flowers and anthers for categorization of anther developmental stages is summarized in Table 1.

The anthers were fixed in 2.5 % glutaraldehyde in 0.1 mol/L phosphate buffer, pH 7.2, placed under a vacuum for 2 h, then stored at 4 °C for several days. After removal from storage, the anthers were rinsed in 0.1 mol/L phosphate buffer for 2 h, and postfixed in 1 % osmium tetroxide overnight. Following post-fixation, anthers were washed in phosphate buffer, dehydrated in an acetone series, embedded in PON-812 resin and cured at 70 °C. Semi-thin sections (1 - 2  $\mu$ m) were cut with glass knives on a LKB-11800 microtome, stained with 0.1% Toluidine Blue, and observed and photographed with a Leica - DM5500B light microscope (Leica, Germany). Ultrathin sections (80 nm) were cut using a Leica-Ultracut S ultramicrotome with a diamond knife, and stained with uranyl acetate and lead citrate. Transmission electron micrographs were taken with a JEOL JEM-1010 (JEOL, Japan) transmission electron microscope at 100 KV.

For SEM, the pollen grains were mounted on copper stubs with a strip of double-sided conductive tape and air-dried. The sample was then coated with gold in a JEOL JFC-1600 sputter coater. Observations and digital images were collected with a JEOL JSM-6360LV SEM. Measurements of the polar axis (P) and equatorial diameter (E) were made on digital SEM images with JEOL's Smile View software.

## 1.2 Sample preparation for RNA extraction

Transcriptome profiling was carried out on developing anthers at four landmark stages: microspore mother cell (MMC) stage, tetrad (TET) stage, uninucleate microspore (UNM) stage and bicellular pollen (BCP) stage. The four stages were defined on the basis of cytological observations with LM and TEM as described above and confirmed using DAPI staining.

Fresh flowers were collected from three well-managed populations of *H. patens* in the SCBG. The anthers were picked out under a dissecting microscope, immediately frozen in liquid nitrogen and stored at -80 °C until RNA extraction. They were classified into four separate groups according to length and each group was pooled by mixing equal quantities of anthers from the three populations.

Total RNA was extracted using pBIOZOL method. RNA quality was characterized by using a NanoDrop ND1000 spectrophotometer (NanoDrop, USA), and by determination of the RIN (RNA Integrity Number) value ( $\geq 7.3$ ) using an Agilent 2100 Bioanalyzer.

## 1.3 Library preparation and sequencing

The cDNA libraries were prepared following manufacturer's instructions (Illumina, USA). Five cDNA libraries (MMC, TET, UNM, BCP and MSA) were established. MSA was a mixed stage anther sample containing equal amounts of



anthers from each of the four stages (MMC, TET, UNM and BCP). mRNA was enriched using oligo (dT) magnetic beads from 5 µg total RNA (total RNA amount of MSA was 20 µg). To avoid priming bias during cDNA synthesis, isolated mRNAs were first fragmented into short pieces (about 200 bp) using RNA Fragmentation Reagents (Ambion, USA). The cleaved mRNA fragments were converted to double-stranded cDNA using random hexamer primers (Illumina) with the SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen, USA). The double-stranded cDNA was purified using a QiaQuick PCR Purification kit (Qiagen, USA) and was processed by end-repair using End Repair Mix Reaction System (Beijing Genomic Institute, China) and the addition of a single adenine. The repaired cDNA fragments were then ligated with sequencing adapters. To select a size range of templates for downstream enrichment, the products of the ligation reaction were purified on a 2% TAE-agarose gel. Then the purified cDNA was enriched by PCR amplification using Primer PE 1.0 and PE 2.0 (Illumina) complementary to the ends of the adapters with Phusion DNA Polymerase. The library products were then sequenced using Illumina HiSeq<sup>TM</sup> 2000 at the Beijing Genomic Institute, China.

#### 1.4 Data processing and *de novo* assembly

The original data produced by the sequencer were defined as raw reads. Base calling accuracy was measured by the Phred quality score (Q score), which is the most common metric used to assess the accuracy of a sequencing platform. Q20 stands for 99 % accuracy representing probability of incorrect base call being 1 in 100, while Q30 stands for 99.9 % accuracy representing probability of incorrect base call being 1 in 1000 (Illumina technique notes: sequencing). Filtering of raw reads was carried out to obtain clean reads by remove those, a, containing adaptors, b, containing more than 5 % unknown nucleotides, and c, showing more than 10 % of bases with a Phred scaled probability (Q) less than 20.

Clean reads were assembled into unigenes using the short reads assembly program Trinity, release-20130225 (<http://trinityrnaseq.sourceforge.net/>). The resulting unigenes were divided into two classes: clusters containing several unigenes that had a similarity of more than 70 %, and a singletons class (Grabherr et al., 2011; Iseli et al., 1999).

#### 1.5 Functional annotation and analysis

To assign putative gene function, unigene sequences were firstly aligned by BLASTx (E-value < 1e-5) to protein databases (NR, SwissProt, KEGG, COG). Then the unigenes were searched against nucleotide database NT (E-value < 1e-5) using BLASTx to retrieve proteins with the highest sequence similarity with the given unigenes and their protein functional annotations. Functional categories of the predicted genes were obtained by applying gene ontology (GO) terms to the NR database annotation using the Blast2GO program. GO functional classification for all unigenes and the distribution of gene functions were analyzed using WEGO software (Conesa et al., 2005; Ye et al., 2006). To identify possible functions, the annotated unigene sequences were searched in the COG database where orthologous gene

products are classified (Kanehisa et al., 2008). To identify active pathways in *H. patens* anthers, annotated sequences were mapped to reference pathways in the KEGG database (Release 63.0). COG and KEGG pathway annotations were performed using BLASTall software.

### 1.6 Comparison with the coffee genome

Since *Coffea canephora* is the only species in the Rubiaceae with a sequenced genome, comparison analysis was conducted between the unigenes of *H. patens* and the *C. canephora* CDSs using BLASTn. Sequence similarity was estimated by the coverage of the matched length, by calculating the value of matched sequence length versus unigene length and CDS length respectively. The coffee genome was downloaded from the following website (<http://coffee-genome.org/download>). Moreover, the expression level and the GO annotations were examined on a set of genes with relatively high sequence similarity to their orthologs in coffee.

### 1.7 Digital Gene Expression Tag Profiling (DGE) analysis

In order to predict putative genes correlating with tapetum and pollen development, four separate DGE libraries (MMC, TET, UNM and BCP) were established as mentioned above and analyzed. Procedures of library construction and sequencing were the same with the MSA library. All clean reads from the four libraries were mapped onto the transcriptome library (MSA) to calculate the expression level for each gene and the gene coverage using SOAP (version 2.21). If there was more than one transcript for a gene, the longest one was used to calculate its expression level and coverage. The gene expression level was calculated using RPKM (Reads Per kb per Million reads) value. The gene coverage was determined as the ratio of the base number in a gene covered by unique mapping reads to the total bases number of the gene. RPKM value  $\geq 10$  was used as a criterion to define high-level expressed genes. RPKM value of 10 is the average of the median for every gene plus 5. Highly expressed genes were categorized with a Venn diagram.

DGE was used to compare the differences in gene expression (Audic and Claverie, 1997). The threshold  $FDR < 0.001$  and the absolute value of the  $\log_2 \text{Ratio} \geq 1$  were used to determine the potential difference in gene expression. Genes enriched at a specific stage were identified by comparing gene expression between two adjacent stages. Genes that were two-fold up-regulated and expressed at high-level (RPKM  $\geq 10$ ) at the same stage were identified. Differentially expressed genes (DEGs) were two-fold up-regulated at only one stage or simultaneously at two stages or three stages, which were identified by pairwise comparison of the four stages of anther development. Among the two fold up-regulated genes at only one stage, those with high expression (RPKM value  $\geq 10$ ) at the same stage and meanwhile with low expression (RPKM value  $\leq 10$ ) at the other three stages were defined as specifically expressed genes (SEGs). Then we carried out GO functional enrichment and KEGG pathway enrichment analysis for these SEGs. Cluster analysis



of gene expression patterns was performed with Cluster (De Hoon et al., 2004) and Java Treeview (Saldanha, 2004) softwares. With GO enrichment analysis, first all DEGs were mapped to GO terms in the database (<http://www.geneontology.org/>), then the gene numbers for every term was calculated, hypergeometric test was used to find significantly enriched GO terms in DEGs. The calculated p-value went through Bonferroni Correction, taking corrected p-value  $\leq 0.05$  as a threshold.

With Nr annotation, Blast2GO program was used to get GO annotation of DEGs, and then WEGO software (Ye et al. 2006) was used to do GO functional classification for DEGs. KEGG pathway enrichment analysis was used to identify significantly enriched metabolic pathways or signal transduction pathways in DEGs (Kanehisa et al., 2008).

### 1.8 Real-time quantitative PCR (q-PCR)

qRT-PCR was performed on independently collected samples to verify the potentially differentially expressed genes (DEGs) identified by RNA-seq. Total RNA was prepared from *H. patens* anthers at the four developmental stages as the same for digital gene expression analysis. First-strand cDNA was synthesized using a FastQuant RT Kit (TIANGEN, China) according to the manufacturer's protocol. Real-time PCR primer was designed on Web Primer: DNA and Purpose Entry website (<http://www.yeastgenome.org/cgi-bin/web-primer>). Primers used in the experiment are listed in Supplementary Table S1. All reactions were performed using GoTaq qPCR Master Mix (Promega, USA). Reactions were carried out in a total volume of 10  $\mu$ L reaction mixture containing 5.0  $\mu$ L of GoTaq qPCR Master Mix (Promega), 0.2  $\mu$ L (10  $\mu$ mol/L) of each primer, 200 ng of template cDNA. The real time RT-PCR amplification was performed with LightCycler 480 II Real-Time PCR System using two-step cycling conditions of 95  $^{\circ}$ C for 10 min, followed by 40 cycles of 95  $^{\circ}$ C for 15s and 60  $^{\circ}$ C for 60s. Dissociation stage condition was set at 95  $^{\circ}$ C for 15s, 60  $^{\circ}$ C for 15s and 95  $^{\circ}$ C for 15s. The  $\beta$ -Actin primer was designed based on the homologous gene of *Rubia cordifolia* L.  $\beta$ -Actin was amplified from *H. patens* anthers at the same four stages and used as an internal control. The relative quantities of transcripts were calculated using the comparative Ct method and three biological replicates were performed.

## 2. Results

### 2.1 Ultrastructural observations of anther wall and pollen development

Anthers of *H. patens* contain four microsporangia. In cross sections, the microspore mother cells (MMCs) are angular in shape and possess a large nucleus with a darkly staining nucleolus (Fig. 1A). The anther wall at the MMC stage is differentiated into epidermis, endothecium, middle layer and tapetum. The tapetum is uninucleate, one or two layered and the microspore mother cells become oval prior to meiosis (Fig. 1A).

After meiosis, tetrads of haploid microspores fill the anther locules (Fig. 1B).

Tetrads are surrounded by a thick asymmetric callose envelope and the primary cell wall is visible outside of the callose. The plasmalemma of the microspores is initially straight and in direct contact with the callose (Fig. 2A). Subsequently, a fibrillar surface coat develops between the callose and the plasmalemma, which has a loose, irregular fibrillar texture and is considered primexine matrix (Fig. 2B). Rod-shaped electron-dense units are radially oriented in the distal part of the primexine, which eventually form the columellae. At this stage, the epidermis, endothecium and middle layer become high vacuolated, nuclei are displaced to the wall and nucleoli are fragmented. The middle layer becomes flat (Fig. 1B). Tapetal cells contain numerous ribosomes and extensive layers of endoplasmic reticulum that lie below the cell membrane and surround the nuclei in concentric rings. Clusters of small vesicles are present throughout the tapetal cytoplasm and pre-Ubisch bodies are formed (Fig. 2C).

At uninucleate stage, microspores are released from the tetrads into the anther locule, and increase both in volume and in wall thickness (Fig. 1C). The epidermis and endothecium are high-vacuolated and contain lipidic materials and starch grains, whereas the middle layer is almost degenerated. Abundant Ubisch bodies are produced in the tapetal cells and extruded through their inner tangential and radial walls (Fig. 2D). The Ubisch bodies possess an irregular and granular shape. The electron density of the Ubisch body wall is very similar to that of the exine. During the early development of free microspores, the columellae increase in thickness, particularly at their distal ends, the tectum develops, while the sporopollenin-like materials are deposited between the distal portions of the columellae. The foot layer occurs at the bases of the columellae (Fig. 2E). White-line-centered lamellae were observed between the foot layer and the endexine (Fig. 2F). The endexine continues to develop on the proximal surface of these white lines and abundant electron-dense material, probably sporopollenin, appears beneath the endexine. Subsequently, the columellae become solid and the intercolumnellar space is filled with fibrillar materials, i.e., remnants of the primexine matrix. As the microspores form a large vacuole (Fig. 2G), the plasmalemma in the interapertural regions recedes, creating a large pericytoplasmic space under the radially oriented membranous granular material layer (Fig. 2H). This event might indicate the start of intine development. At this phase, the tapetal cell walls and cytomembranes are completely degenerated, a large number of vesicles and lipids are present throughout the tapetal cells and Ubisch bodies are released from the surface of the tapetal cells into the locule (Fig. 2I). At this phase, a thin layer of cuticle is present at the anther epidermis (Fig. 2I).

Initiation of intine development takes place prior to the first pollen mitosis. At this stage the intine has a fibrillar structure and microspores possess abundant organelles e.g., rough endoplasmic reticulum (RER) and a large vacuole (Fig. 3A). The cytoplasm contains long profiles of rough endoplasmic reticulum parallel with the plasmalemma and numerous mitochondria. Later the microspores enter bicellular pollen stage in which the generative cell is surrounded by lipid droplets (Fig. 3B). The cytoplasm is characterized by an abundance of compound starch grains and lipid

droplets (Fig. 3B). The intine thickens and appears more electron-dense and reaches the endexine at several sites (Fig. 3C). Initiation of the characteristic fibrous thickenings takes place in the endothecium (Fig. 1D). Cells of the tapetum and middle layer completely degenerate, leaving the endothecium as the innermost anther wall layer (Fig. 3D). The volume of the pollen grains progressively increases, resulting in stretching of the pollen wall, which is more pronounced in pollen grains at dehiscence. Projection of the intine in the aperture region was observed in some pollen grains (Fig. 3D, F). Pollen grains are bicellular at anthesis, relatively small [P (polar axis) 18.36 (17.8-19.4)  $\mu\text{m}$   $\times$  E (equatorial diameter) 19.66 (18.7-20.9)  $\mu\text{m}$ ], and oblate spheroidal in equatorial view (Fig. 3E). The grains are planaperturate with three compound apertures. The exine pattern is micro-reticulate with the sculpting somewhat obscured by the presence of fibrillar material that fills the spaces between the columellae (Fig. 3F).

## 2.2 cDNA sequence generation and *de novo* assembly

Approximately 948 million bases were generated and a set of 105 million 90 bp paired-end reads were obtained after cleaning and quality checks. Q20 and Q30 values reached 97.8 % and 92.1 % respectively, representing high quality sequencing. Assembly of reads resulted in 177,548 contigs with an average length of 354 nt. These were assembled into 89,849 unigenes, which were divided into two classes. The first class is distinct singletons (52,379) with the prefix unigene, representing unigenes derived from a single gene. The second class is distinct cluster unigenes with the prefix CL and the cluster ID, where each cluster contains unigenes with similarities of more than 70 %. The 37,470 cluster unigenes are distributed into 14,948 distinct clusters, among which, 3,263 contain one unigene (unigenes shorter than 200 bp were removed after family clustering), and 7,247 contain two unigenes. The remaining 4,438 clusters contain three or more unigenes, and the largest two clusters contain 45 unigenes. The cluster unigenes may be derived from the same gene or from a homologous gene (Grabherr et al., 2011).

## 2.3 Gene annotation and functional classification

In the whole unigene set, a total of 57,476 (64 %) unigenes were significantly matched to known genes in the public databases of NR, NT, Swiss-Prot, KEGG, COG and GO (Table 2), representing putative functional annotations for more than half of the assembled sequences.

Homology searches performed using BLASTx against the non-redundant (NR) protein database showed 55,563 (61.8 %) of unigenes with matches to the NR database with an E-value cut-off of  $1\text{e-}5$  (Supplementary Fig. S1A). The similarity distribution shows 68.6 % of BLASTx hits are within the range 60 to 100 %, representing strong homology. Only 31.4 % have similarity values less than 60 %, showing moderate homology (Supplementary Fig. S1B). With regard to species similarity, the highest proportion of matched sequences in the NR database are derived from *Solanum lycopersicon* (33.9 %, 18,815 unigenes) and *Vitis vinifera*

(26.4 %, 14,667 unigenes) which belong to the Solanales, a sister order to the Gentianales, to which Rubiaceae belongs (APG III, 2009; Supplementary Fig. S1C). *H. patens* shows a similar proportion (27.2 %, 15,090 unigenes) of sequences matching the genome of *Coffea canephora*, which was published recently (Denoëud et al., 2014; Dereeper et al., 2015). Further comparative analysis between *H. patens* and *C. canephora* is presented in section 2.4 below.

To understand the potential function of the assembled unigenes, the COGs (clusters of orthologous group of proteins) functional annotation system was used. In total, 18,948 (21.1 %) unigenes were mapped to the COG database and possible functions and statistics were predicted. The five largest categories are general function (17.9 %), transcription (8.9 %), replication, recombination and repair (8.6 %), signal transduction mechanisms (7.0 %), and posttranslational modification, protein turnover and chaperones (6.9 %) (Supplementary Fig. S2). The proportion of genes annotated by COG and the predicted categories are quite similar in other studies of non-model species without known genomic sequences in the past a few years; for instance, 20.4 % in *Triadica sebifera* in Euphorbiaceae (Divi et al., 2016), 20.1 % in *Uncaria rhynchophylla* in Rubiaceae (Guo et al., 2014) and 21.1% in *Dendrocalamus latiflorus* in Gramineae (Zhang et al., 2012).

GO assignments were used to classify the functions of predicted genes, resulting in 42,244 (47.0 %) unigenes assigned to at least one GO term, while 47,605 (53.0 %) could not be assigned. Among the 42,244 BLASTable (E-value < 1e-5) unigenes, 17,923 unigenes are distinct singletons, while 24,321 are distinct cluster unigenes, distributed in 9,639 clusters. If the unigenes contained in the distinct clusters were treated as single unigenes, then the total number of annotated unigenes would be 27,562, and the annotated proportion would be 41.0 %, which is close to the proportion of unigenes in the same cluster treated as independent unigenes. Non-BLASTable (E-value > 1e-5) sequences have been reported in all studied plant transcriptomes, with the proportion varying from 13.0 % to 80.0 %, depending on the species, the sequencing depth and the parameters of the BLAST search (Parchman et al., 2010; Blanca et al., 2011; Ness et al., 2011; Zhang et al., 2012; Zhang F.J. et al., 2014). Non-BLASTable (E-value > 1e-5) sequences might result from biological factors, including rapidly evolved genes with divergent sequences, species-specific genes and the persistence of non-coding fractions derived mainly from untranslated regions (Zhang et al., 2012; Logacheva et al., 2011). The BLASTable (E-value ≤ 1e-5) unigenes are divided into three GO categories: biological process (32,700), cellular component (33,751), and molecular function (31,293). In the biological process category, the most abundant sequences are classified into cellular process (81.0 %), metabolic process (77.0 %) and single-organism process (58.4 %). In the cellular component category, cell (93.6 %), cell part (93.6 %), and organelle (74.6 %) are the three most represented GO terms. In the molecular function category, catalytic activity (66.7 %), binding (63.6 %), and transporter activity (10.0 %) are highly represented (Supplementary Fig. S3). The constitution of the most represented GO

terms shows a high degree of similarity with the GO annotation of *C. canephora* (Dereeper et al., 2015).

To identify biological pathways activated during anther development of *H. patens*, the assembled unigenes were annotated against the KEGG database (E-value  $< 1e-5$ ). A total of 31,169 (34.7 %) unigenes were mapped into 128 pathways in the KEGG database (Supplementary Table S2). The maps with the highest unigene representation are metabolic pathways (6,975, 22.4 %), followed by biosynthesis of secondary metabolites (3,465, 11.1 %), plant-pathogen interaction (1,796, 5.8 %) and plant hormone signal transduction (1,609, 5.2 %). This constitution exhibits a high degree of similarity with rice anther transcriptome profiles derived from microarray analysis (Deveshwar et al., 2011).

## 2.4 Comparison with the coffee genome

Alignments of the *H. patens* unigenes (89,849) with the coffee genome (25,574 CDSs) revealed 15,090 (16.8 %) unigenes with a significant overlap to 7,313 (28.6 %) coffee CDSs. This proportion is lower than expected given that both species are from the same family. When clean reads (105 million) were mapped to the coffee genome allowing a maximum 2 bp mismatch, a total of 366,227 (0.35 %) reads matched 4,047 (15.8 %) coffee CDSs, a lower proportion than observed for unigene alignments. Meanwhile, the number of reads with no mismatches was 8,821, corresponding to 286 coffee CDSs, and 79,220 reads with a 1 bp mismatch corresponding to 1,549 coffee CDSs. It is predictable that as mismatched bases increase, more matched reads and coffee CDS will result. Read alignment is presumed to be more accurate, since the reads represent the most comprehensive data. The low proportion of matches between *H. patens* and *C. canephora* might be explained by the fact that these species are not genetically close and are members of distant tribes in different subfamilies.

Among the 7,313 coffee CDSs, 3,774 have just one matching unigene, while the remaining 3,539 match multiple unigenes. In these cases, most of those matching unigenes are from the same cluster and are longer than the matched coffee CDS. This is indicative of sequence similarities other than matches with different portions of the same coffee CDS. Among these aligned unigenes of *H. patens*, 80 (0.5 %) are related to anther wall and pollen wall development. The remaining unigenes are largely annotated with response to salt stress, DNA-dependent regulation of transcription, oxidation-reduction process, response to cadmium and protein phosphorylation, which represent general functions (Supplementary Table S3).

## 2.5 Digital Gene Expression Tag Profiling (DGE)

The numbers of unigenes mapped by clean reads in each of the four libraries are 70,735 (MMC), 71,510 (TET), 71,248 (UNM) and 70,288 (BCP). Among these, 29,002 (41.0 %), 28,660 (40.8 %), 26,806 (37.6 %) and 26,920 (38.3 %) genes have RPKM values  $\geq 5$  in MMC, TET, UNM and BCP, respectively. Detecting low



abundance transcripts is one of the advantages of RNA-seq. In this study, 14,059 (19.9 %), 13,731 (19.2 %), 15,399 (21.6 %) and 15,617 (22.2 %) genes are low abundance in MMC, TET, UNM and BCP respectively with  $\text{RPKM} \leq 1$ . Genes with  $\text{RPKM} \geq 10$  at one or more stages were considered highly expressed and the numbers detected at MMC, TET, UNM and BCP stages were 18,593 (26.3 %), 18,129 (25.4 %), 16,995 (23.9 %) and 17,370 (24.7 %) respectively (Fig. 4). It is supposed that unigenes with higher expression are more likely to be full-length, therefore further analyses was focused on genes with  $\text{RPKM}$  values  $\geq 10$  at one or more stages. These are classified into four main categories and 15 sub-categories based on expression pattern. The four main categories are highly expressed ( $\text{RPKM}$  value  $\geq 10$ ) in one, two, three or in all four stages. The numbers of genes of each sub-category are calculated and illustrated in the Venn diagram (Fig. 5). In total, 11,570 genes are highly expressed across all stages, which may suggest their involvement in housekeeping functions or general metabolism.

### 2.5.1 Identification of up and down-regulated genes and their GO annotations

The comparisons of adjacent stages, identified 2,851, 2,491 and 4,781 genes that were potentially up-regulated by at least two-fold, while 3,443, 3,591 and 4,305 genes were potentially down-regulated in TET, UNM and BCP stage, respectively. A greater number of genes were down-regulated in comparison to those up-regulated in TET and UNM, however, this trend is reversed in BCP where a larger proportion of genes showed up-regulation (Fig. 6). Among these putatively up-regulated genes, those with high expression ( $\text{RPKM}$  value  $\geq 10$ ) in one particular stage and lower expression ( $\text{RPKM} < 10$ ) in the other three stages, were considered stage-enriched. 927 (32.5 %), 454 (18.2 %) and 2,068 (43.3 %) genes are enriched in TET, UNM and BCP, respectively (Fig. 6). BCP has the largest proportion of stage-enriched genes in all the stages analyzed, moreover, both up and down-regulated genes increase dramatically in BCP in comparison to earlier stages, highlighting the distinct gene expression profile at this stage.

Functional association of the potentially up-regulated and stage-enriched genes based on Gene Ontology (GO) annotations (Supplementary Table S4) revealed that TET stage-enriched and up-regulated genes were enriched in protein phosphorylation, metabolic process, oxidation-reduction process, histone H3-K9 methylation, DNA methylation, DNA-dependent regulation of transcription. UNM stage-enriched and up-regulated genes were enriched in oxidation-reduction process, DNA-dependent regulation of transcription, proteolysis, metabolic process, transmembrane transport and response to cadmium ion. Categorization of BCP stage-enriched and up-regulated genes into GO functional groups showed enrichment for genes involved in oxidation-reduction process, pollen tube growth, plant-type cell wall modification, protein phosphorylation and transmembrane transport, which could be important contributors to pollen maturation and the pollen transcriptome. However, a large number (38,440) of BCP expressed genes have not been previously reported, which could serve as a useful resource to mine transcripts for validation of putative



gametophyte and/or germline functions.

Genes with GO terms (Supplementary Table S5) related to anther wall and pollen wall development were further selected (Fig. S4). Among the putatively up-regulated genes in TET stage, 40 (46.5 %) are related to pollen exine development, 40 (46.5 %) are related to other components of, or contributing to, pollen wall development and 6 (7.0 %) are related to tapetum development. Among the stage-enriched genes in TET stage, 6 (35.3 %) are related to pollen exine development, 9 (52.9 %) to pollen wall development and 2 (11.8 %) to tapetum development. At the UNM stage, 11 (29.0 %) putatively up-regulated genes are related to pollen exine development and 27 (71.0 %) are related to other components of, or contributing to, pollen wall development, but none to tapetum development. In addition, there is 1 stage-enriched gene (Unigene6816) related to pollen exine development and 1 (Unigene9576) contributing to hemicellulose metabolic process, but none contributing to tapetum development. At the BCP stage, 50 (21.4 %) putatively up-regulated genes are related to pollen exine development, 180 (76.9 %) to other components of or contributing to pollen wall development and four (1.7 %) to tapetum development. At BCP stage, 20 (24.1 %) stage-enriched genes are related to pollen exine development, 63 (75.9 %) to other components of or contributing to pollen wall development and none related to tapetum development. From TET stage onwards, the genes related to pollen wall development first decrease in UNM then increase in BCP, while the genes related to tapetum development decrease in both UNM and BCP. This trend is consistent with developmental events in pollen wall and tapetum.

### 2.5.2 Identification of putatively stage-specific expressed genes and their expression profiles

Pairwise comparisons within the four DGE libraries identified 7,720 genes are differentially expressed across the four stages. These genes were categorized into 14 groups based on up or down regulation. Hierarchical cluster analysis was performed on each of the groups to illustrate the results (Fig. 7).

In group 1 to 4, the genes are at least two-fold up-regulated in one particular stage comparing to any of the other three stages and the differences in expression level are not significant. In total, 1,644 (21.3 %), 1,637 (21.2 %), 688 (8.9 %) and 3,220 (41.7 %) genes illustrated in the groups 1 to 4 show expression peaks in MMC, TET, UNM and BCP, respectively. Gene Ontology (GO) annotation analysis of the differentially expressed genes (Supplementary Table S6) indicated that a large number of genes were involved in oxidation-reduction processes, protein phosphorylation, pollen tube growth, response to salt stress and response to cadmium ion. A total number of 86 genes (1.1 %) in the groups 5 and 10 show two-fold up-regulation in the former or the latter two stages, coinciding with the time of tapetum formation or tapetum PCD. These genes participate in oxidation-reduction processes, metabolic processes and DNA-dependent regulation of transcription. Genes in group 7 are down-regulated at tetrad and uninucleate microspore stages, while genes in group 8 are two-fold up-regulated at these two stages. The genes in both groups were mainly

enriched for protein ubiquitination. Genes in group 6 are down-regulated at tetrad and bicellular pollen stages, whereas group 9 genes are two-fold up-regulated at these stages. Group 9 genes are involved in metabolic processes, sodium ion transmembrane transport, regulation of pH, pollen tube growth and plant-type cell wall modification. In groups 11 to 14, a total of 427 (5.5 %) genes are two-fold down-regulated in one particular stage. The genes in these 4 groups were mainly related to oxidation-reduction process, ATP catabolic process, translation, DNA-dependent regulation of transcription. In total 3,427 (44.4 %) genes in all 14 groups were not annotated by GO terms and as such, deserve further attention as a source of genes with unidentified roles in anthers.

Groups 1 to 4 were further analyzed to identify stage specifically enriched genes. They were defined as two-fold up-regulated and highly expressed ( $\text{RPKM} \geq 10$ ) in one particular stage, while expression values in the other three stages were less than 10. As a result, 3,769 stage-specifically enriched genes across the four stages were obtained. The numbers at each stage were, 776 (20.6 %), 807 (21.4 %), 322 (8.5 %) and 1,864 (49.5 %) at MMC, TET, UNM and BCP respectively (Fig. 8). UNM had the lowest share of stage specifically enriched genes, while BCP had the largest share. Thus, BCP showed the most diverse expression profiles compared to the other three stages.

The expression patterns of all stage specifically expressed genes were analyzed and genes with similar expression profiles were clustered. A total of 14 clusters were generated (Fig. 9). Genes in the same cluster are considered co-expressed and may be targets of the same transcription factors. The expression profile of the largest cluster (“a”; 1,487 genes) showed up-regulation in MMC and TET followed by down-regulation in UNM and BCP, coinciding with the pattern of tapetum development. Further, GO terms relating to tapetum development are strongly enriched in cluster “a” (Supplementary Table S7) confirming the cytological changes of tapetum development from MMC to BCP involve gene down-regulation. Clusters “b” (242) and “g” (132) also exhibit down-regulation from MMC to BCP and are dominated by genes related to oxidation-reduction process, DNA-dependent regulation of transcription and metabolic process. Clusters “j” (246), “m” (237) and “n” (375), show up-regulation trends from MMC to BCP and are dominated by genes related to oxidation-reduction process, pollen tube growth, plant-type cell wall modification and cellular membrane fusion. In clusters “d” (102), “c” (95) and “f” (90), genes were up-regulated at both or either of the TET and UNM stages and annotated by lipid metabolic process, very long-chain fatty acid metabolic process, transmembrane transport and intracellular signal transduction. These features coincide with the finding that tapetal cells are most active during TET and UNM stages and are known to be involved in the synthesis of flavonoids and other secondary metabolites that eventually are transported to developing microspores. In clusters “i” (135), “l” (199) and “h” (332), genes are down-regulated at both or either of the TET and UNM stages and annotated by pollen tube growth, oxidation-reduction process, plant-type

cell wall modification and negative regulation of programmed cell death. Genes in cluster “k” (66), showing up-regulation at TET and BCP stages, are mainly involved in xylan biosynthetic process, glucuronoxylan metabolic process, plant-type cell wall modification, cellulose biosynthetic process and plant-type cell wall biogenesis. Coincidentally, callose wall, primexine and intine related to polysaccharide metabolism are synthesized mainly at TET and BCP stages. Genes down-regulated at TET and BCP stages are clustered in “e” (31), and are enriched in ATP catabolic process, metabolic process, transmembrane transport and DNA-dependent regulation of transcription.

### 2.5.3 Candidate genes relevant to tapetum and pollen wall development

Differentially expressed genes and specifically expressed genes enriched for the relevant GO terms of anther wall and pollen wall development were identified. In total 243 differentially expressed genes and 108 stage-specific genes were obtained (Fig. S5). Genes related to sporopollenin biosynthetic processes, are only enriched at TET (6) and UNM (1) stages. Genes related to pollen exine formation, are mainly enriched at TET (26 genes) and BCP (35 genes) stages, among which 5 (19.2 %) and 19 (54.3 %) genes are specifically expressed. Only 1 gene related to tapetal layer morphogenesis is specifically expressed at TET stage. There are 4 and 3 genes related to tapetal layer development differentially expressed at MMC and TET stages, respectively. Among those genes, 3 are specifically expressed at MMC and 1 is specifically expressed at TET stage. Genes relating to pectin are mainly enriched at MMC and BCP stages, with 14 (MMC) and 26 (BCP) genes being differentially expressed and 11 of which are specifically expressed at BCP stage. Genes relating to cellulose mainly express at MMC, TET and BCP stages, with the numbers of differentially expressed genes being 15, 12 and 41, and the numbers of specifically expressed genes as 5, 4 and 20 respectively. Genes corresponding to hemicellulose metabolic processes are only expressed at TET, UNM and BCP stages, whilst 88.4 % (38 genes) of these are differentially expressed at BCP stage and among these genes, 30 are specifically expressed. These findings are consistent with the cytological changes and developmental events in anther wall and pollen wall development.

Pollen wall development requires lipid and polysaccharide metabolism, therefore genes enriched in carbohydrate and lipid metabolism pathways were analyzed (Fig. S6). The numbers of differentially expressed genes in both pathways only slightly change between MMC and UNM stages, while a marked increase occurs at BCP stage. This trend emphasizes the distinct gene expression profile at BCP stage. At MMC and UNM stages, a greater number of stage-enriched genes are involved in lipid metabolism in comparison to those involved in carbohydrate pathways (Supplementary Table S8). At TET and BCP stages, this trend is reversed, where fewer stage-enriched genes are involved in lipid metabolism than in carbohydrate metabolism. The MMC stage is enriched with a greater number of genes relating to glycerophospholipid and lipid metabolism, while TET stage is enriched in starch and sucrose metabolism as well as glycerophospholipid metabolism. UNM is enriched

with the greatest number of genes relating to cutin, suberine and wax biosynthesis. The first three largest numbers of genes at BCP are related to starch and sucrose metabolism, pentose and glucuronate interconversions, and glycerophospholipid metabolism. This might suggest nutrient synthesis and storage at BCP stage, consistent with the observation that the bicellular pollen cytoplasm is characterized by an abundance of starch grains and lipid droplets.

## 2.6 Verification of expression profiles of selected unigenes related to tapetum and pollen wall development by qRT-PCR

Real-time quantitative PCR (qRT-PCR) analysis was carried out on 20 candidate genes selected at random (Fig. 10). Among these, 14 genes show high expression with  $RPKM \geq 10$  in at least one stage, while 6 display low expression with  $RPKM \leq 10$  at all four stages. The expression values at developmental stages derived from RNA-seq and q-PCR of each gene are listed in Supplementary Table S9. The expression patterns of all 20 genes analyzed by qRT-PCR largely agree with the digital gene expression tag profiling (DGE), as the correlation co-efficients ( $r$ ) are all greater than 0.9. This result provides support for the reliability of RNA-Seq data and the differential gene expression profiles observed for anther stages of *H. patens*.

## 3. Discussion

### 3.1 Tapetum and pollen development display common features in Rubiaceae

Anther development is a complex process which involves transition from sporophyte to gametophyte, control of mitotic and meiotic cell divisions, together with the coordination of pollen and anther maturation (McCormick, 1993; Scott et al., 2004; Zhang et al., 2011). Ultrastructural changes involved in anther development have been described for many angiosperms, especially for model species including *Arabidopsis* and rice (Sanders et al., 1999; Zhang and Wilson, 2009; Zhang et al., 2011), although there are differences in the resolution and detail of the stages described (Owen and Makaroff, 1995). In our study, anther and pollen development of *H. patens* was shown to be typical of many angiosperms, with some features common in the Rubiaceae family. Our observations of developmental events and cytological changes provides a solid foundation for the analysis of gene expression profiles in the developing anthers of *H. patens*.

The tapetum cells in *H. patens* maintain integrity and position indicating that the tapetum is of the secretory type, which fits the description of a type five tapetum (Pacini, 1997), the most common in angiosperms (Owen and Makaroff, 1995). Ubisch bodies occurred simultaneously with the developing pollen exine, which corroborates the idea that they are required for transferring tapetum-derived sporopollenin precursors to the exine (Huysmans et al., 1998). The tapeta of *Arabidopsis* and rice are also secretory but use different export routes for sporopollenin. In *Arabidopsis*, specialized organelles, elaioplasts and tapetosomes, are supposed to be transporters. Lipids are transported from the tapetum to microspores in two ways: vesicular transport and use of transporters as lipid carriers (Vizcay-Barrena and Wilson, 2006).

In rice and other cereals, Ubisch bodies are thought to export sporopollenin across the hydrophilic cell wall to the locule (Zhang et al., 2011). Thus, the tapetum in dicots and monocots might have different mechanisms for sporopollenin translocation. Analogies are often found between the ornamentation of the pollen exine and that of the Ubisch body wall (Hesse, 1986; Vinckier and Smets, 2002, 2003). Ubisch bodies are considered have great potential as model system to study sporopollenin deposition, since they are acellular structures, independent of cytoplasmic control (Clément and Audran, 1993; Verstraete et al., 2014). To date, only the *RAFTIN* gene was identified in pro-orbicle bodies and shown to accumulate in Ubisch bodies in wheat and rice. *RAFTIN* is highly anther-specific and essential for pollen development in cereals (Wang et al., 2003). Future genetic studies to identify the genes and gene products that control sporopollenin polymerisation and exine patterning will need to take Ubisch bodies into account when screening for phenotypes (Verstraete et al., 2014).

The fundamental structure of pollen wall in angiosperms consisting of the outer exine and the inner intine is generally conserved (Heslop-Harrison, 1968; Blackmore et al., 2007; Ariizumi and Toriyama, 2011). The components contributing to the pollen wall are produced and accumulated in a precise temporal sequence. Sporopollenin makes up the majority of the material of the exine (Blackmore et al., 2007). In *H. patens*, the sporopollenin deposition starts from TET stage. Intine development starts from the late UNM stage and is complete by the end of BCP stage. The ontogenetic sequence of pollen wall development follows the basic scheme in the family (Hansson and El-Ghazaly, 2000; El-Ghazaly et al., 2001). Intinous projections occurred in the aperture region of *H. patens* pollen grains, known as protruding onci or pollen buds in the Rubiaceae, was believed to be a relatively common feature in this family (Dessein et al., 2005; Kuang et al., 2012).

Besides exine and intine, the callose wall and primexine are successively synthesized and degraded at precise times. The callose wall, consisting of linear  $\beta$ -1,3-glucan polymers (Hird et al., 1993), begins to deposit at the beginning of TET stage and degrades at the end of TET stage in *H. patens*. The primexine, which acts as a template that guides the accumulation of sporopollenin, is composed of neutral and acidic polysaccharides, cellulose and some proteins (Heslop-Harrison, 1968; Scott et al., 2004; Jiang et al., 2013). Primexine matrix usually with fibrillary texture appears between the callose and the plasmalemma at the TET stage in *H. patens*. In addition, lipidic material, which may be classed as pollenkitt (El-Ghazaly et al., 2001; Edlund et al., 2004), is deposited onto the exine surface at the late BCP stage. These ontogenetical features were also described in the available ultrastructural studies in Rubiaceae (Hansson and El-Ghazaly, 2000; El-Ghazaly et al., 2001; Kuang and Liao, 2015).

### 3.2 DEGs expression patterns agree with the developmental events

Genes related to sporopollenin biosynthetic process were closely associated



with fatty acid metabolism and only expressed at TET and UNM stages, consistent with the appearance of pro-Ubisch bodies and Ubisch bodies in tapetal cells. Genes related to pollen exine formation are enriched at TET and BCP stages, which may correspond to the time that primexine formed and exine deposited. Genes related to hemicellulose metabolic process, are mainly enriched at BCP stage, a few are enriched at TET and UNM stages and none are expressed at MMC stage, coinciding with the stages at which the primary cell wall and intine develop. Genes related to pectin are largely enriched at BCP stage, and genes related to cellulose are largely enriched at TET and BCP stages. These gene expression patterns agree well with the observed stages in which intine and primexine appear in *H. patens*. Genes related to tapetal layer morphogenesis and development are enriched at MMC and TET stages, consistent with stages during which tapetal cells are differentiated and pro-Ubisch bodies are produced.

The numbers of genes enriched in different pathways are correlated with structures and components present at different anther developmental stages. At UNM stage, abundant Ubisch bodies are synthesized, which involve lipid metabolism. Consistently, more genes are revealed in lipid metabolism pathways than in carbohydrate metabolism pathways at the same stage. The synthesis and degradation of callose wall and primexine, and the formation of intine are involved in carbohydrate metabolic pathways (Jiang et al., 2013). The genes related to the carbohydrate metabolism are enriched at the stages during which callose wall, primexine and intine form. In addition, the cytoplasm of mature *H. patens* pollen grains is characterized by an abundance of compound starch grains and lipid droplets at BCP stage. This may account for the increase in the number of genes enriched in both carbohydrate and lipid metabolism pathways at this stage.

GO enrichment analysis of the stage-enriched genes and differentially expressed genes suggested the gene number, gene function and expression profile were significantly correlated with anther developmental processes in *H. patens*. The genes filtered from KEGG pathways analysis and GO enrichment analysis, provide a useful resource to identify novel genes and their possible functions in pollen wall development.

### **3.3 Comparative analysis of anther and pollen transcriptomes confirms inter-specific conserved patterns**

Developmental transcriptomic analysis of anthers and pollen grains utilizing microarray platforms have been reported in *Arabidopsis* (Honys and Twell, 2004; Pina et al., 2005), rice (Huang et al., 2009; Fujita et al., 2010; Deveshwar et al., 2011), maize (Ma et al., 2008), and *Brassica rapa* (Dong et al., 2013). Our study revealed 89,849 unigenes in the *H. patens* anther transcriptome, well above the numbers of genes predicted to be expressed in anthers of other species. However, high-throughput sequencing techniques increase the chance of detecting transcripts with low abundance and microarrays have been reported to be less effective at identifying low



abundance transcripts (Loraine et al., 2013). Further, the number of *H. patens* unigenes is an estimate, since the genome is unknown and unigenes sharing sequence similarities were classified as different unigenes, in consideration of alternative splicing events. High numbers of unigenes have also been reported for species with no reference genomes such as for the fruit transcriptome of *Triadica sebifera* with 92,550 unigenes (Divi et al. 2016), the embryo transcriptome of chrysanthemum with 116,697 unigenes (Zhang F.J. et al., 2014) and the leaf transcriptome of *Lycium chinense* with 61,595 unigenes (Wang et al. 2015). Finally, each developmental stage of *H. patens* anther contains about 20 % of transcripts with  $RPKM \leq 1$ , which most likely would not be detected with microarrays. These low abundance transcripts provide a resource for future analyses.

Transcriptomic analysis in developing and germinated pollen of rice displayed a “U-type” change in the number with the lowest number at the bicellular pollen stage (Wei et al., 2010). Developmental analysis of the Arabidopsis pollen transcriptome revealed a major shift in mRNA populations between BCP and TCP stages, reflective of the transition from earlier cell division to later pollen maturity (Honys and Twell, 2004). This distinct phase shift suggests inter-specific conservation of pollen-expressed genes. In *H. patens* anthers, BCP stage has the greatest number of stage-specific genes. The divergence between *H. patens* and the two model species might result from the relative contribution of sporophytic and gametophytic cell types and earlier developmental stages (MMC and TET) in *H. patens* together with taxon-specific features, such as insect-pollination and bicellular pollen at maturity. Moreover, different platforms, tissue collections and comparison methods can result in significant variation in anther transcriptome profiles for different species (Hollender et al., 2014).

Since sample staging of anther development is similar between *H. patens* and rice in the study of Deveshwar et al (2011), additional comparative analysis was carried out. Transcriptome profiling of rice was investigated in anthers at pre-meiotic (PMA), meiotic (MA), single-celled microspore (SCP) and tri-nucleate pollen (TPA) stages (Deveshwar et al., 2011). At PMA and MA stages (MMC in *H. patens*), there was high representation of sporophytic tissue compared to gametophytic tissue. This characteristic coincided with the phenomenon that most of the transcriptome changes corresponding to the sporogenous tissue and developing tapetum. TPA anthers contained a relatively higher cellular mass of gametophytic tissue (pollen), meanwhile prominent differences were found in transcript composition comparing with transcriptomes of earlier anther stages. In *H. patens*, BCP stage anthers were the most distinctive of the four stages, with the highest number of stage-specific and differentially-expressed transcripts (Fig. 8), showing great similarity to the TPA stage in rice (Deveshwar et al., 2011). This common shift, from free microspore-stage anthers to late stage anthers, could be explained by the distinctive transcriptomes of developing pollen and sperm. It also reflects the large number of BCP enriched genes in *H. patens* associated with pollen development and maturation. Comparison of gene expression between two adjacent stages of anther development further highlighted this

major switch of gene expression. The apparent transition from the free microspore stage to the BCP stage in *H. patens* also reflected the switch from the sporophytic to the gametophytic programme (Deveshwar et al., 2011).

### 3.4 Homologs of anther and pollen-expressed transcription factors were revealed

Putative homologs of known transcription factors expressed in developing anthers were identified in *H. patens* (Supplementary Table S10). The expression patterns of these genes were largely consistent with those of their counterparts in other species, supporting the conservation of regulatory features in anther and pollen development in *H. patens*.

In the early stages of Arabidopsis anther development, DYT1 is required for tapetum development and function (Zhang et al., 2006), directly regulating *TDF1* expression in the tapetum (Feng et al., 2012). The highest expression of both *DYT1* and *TDF1* occurred at similar stages of tapetal development (Gu et al., 2014). CL1502.contig1 of *Hamelia patens* shares high amino acid sequence identity with *DYT1* in the bHLH domain. Unigene29472 is a putative homolog of *TDF1*. Both CL1502.contig1 and Unigene29472 show expression peaks at MMC stage, weak expression at TET or/and UNM stages, but no expression at BCP stage. *AtMYB103*, another important transcription factor for tapetum function and male fertility, is also regulated by DYT1 (Feng et al., 2012). A candidate homolog of *AtMYB103*, CL6214.contig1, shows peak expression at MMC stage and lower expression at TET stage, but no expression at UNM and BCP stages. The expression pattern of these three putative homologs correspond to the early stages of tapetum development and callose dissolution, and exhibit somewhat consistent patterns with those of *DYT1*, *AtMYB103* and *TDF1*, which suggests shared pathways in the regulatory network controlled by DYT1 in Arabidopsis and *H. patens*.

CL4992.contig3 shares high amino acid sequence identity with the MYB domains of *AtMYB97*, *AtMYB101* and *AtMYB120*. It shows expression peak at BCP stage, indicating a consistent profile with these three MYB genes. DUO POLLEN1 (DUO1) is an *Arabidopsis* male germline-specific R2R3-type MYB transcription factor (Rotman et al., 2005; Durbarry et al., 2005) that is essential for male germ cell division and differentiation (Brownfield et al. 2009). CL1108.Contig1, homologous to *AtDUO1*, shows peak expression at BCP stage, consistent with that of *AtDUO1*. It is likely that some of the pathways regulated by MYB factors are shared by *Arabidopsis* and *H. patens*.

The AtMIKC\* transcription factors (TFs) AGAMOUS-LIKE30 (AGL30), AGL65, AGL66, AGL94 and AGL104 are mainly expressed in mature pollen and play an essential role in transcriptional regulation during late pollen development (Verelst et al., 2007a; Verelst et al., 2007b). In *H. patens*, Unigene32598 and Unigene1629 share high amino acid sequence identity in the MIKC\* domain, representing probable homologs of MIKC\* transcription factors. Both of these are predominantly expressed at BCP stage, consistent with the expression patterns of AtMIKC\* genes, which indicates conserved regulatory features of late pollen development and pollen maturation in *H. patens*.

### 3.5 Expression profiles of homologs with verified conserved patterns in pollen development

The *H. patens* homologs of known genes with characterized functions in exine formation in Arabidopsis or rice were identified (Supplementary Table S10). Their expression profiles were analysed and verified by Q-PCR. *TKPR1*, which is involved in a conserved biosynthetic pathway in sporopollenin monomer biosynthesis, shows peak expression at tetrad stage (Grienenberger et al., 2010). Unigene32349, a homolog of *TKPR1*, is putatively highly expressed at TET stage, consistent with the expression profile of *TKPR1*. *LAP6* and *LAP5* are specifically expressed during the period of exine synthesis and are essential for the exine production (Dobritsa et al., 2010). The homologous unigenes CL12695Contig1 and CL3181Contig1 show peak expression at TET and UNM stages. *LAP3* is essential for pollen development and proper exine formation (Dobritsa et al., 2009), and the homologous Unigene27400 shows peak expression at TET stage. Unigene34182 shows peak expression at TET and UNM stages and represents the *H. patens* homolog of *MS2*, which is involved in exine development and is expressed in tapetum shortly after microspore release (Aarts et al., 1997, Gómez et al. 2015). Unigene33351 and CL9729-Contig3 show higher read counts at TET stage and are homologous to the gene *ABCG26 / WBC27*. *ABCG26* is expressed specifically in tapetal cells at the early vacuolate stage and plays a crucial role in the transfer of sporopollenin lipid precursors from tapetal cells to anther locules (Choi et al., 2011), while *WBC27* is expressed during early stages of anther development (Xu et al., 2010).

Unigenes homologous to known genes related to tapetum development, intine development and callose synthesis were also identified in *H. patens* (Supplementary Table S10). Unigene30846, homologous to *AtUSP* (Schnurr et al., 2006), is suggested to be involved in intine development and putatively highly expressed at BCP stage, consistent with the period of intine accumulation. Moreover, Unigene30846 matches Cc10\_g13640 of coffee and shares the functional GO annotation of UDP-sugar pyrophosphorylase, which suggests functional similarity. Unigene29363 is homologous to *CalS5*, which encodes a callose synthase (Dong et al., 2005). The annotation of Unigene29363 (0052543 // callose deposition in cell wall, 0006075 // (1->3)-beta-D-glucan biosynthetic process and 0009556 // microsporogenesis) and peak expression at TET stage are consistent with the function and expression pattern of *CalS5*. *RTS2*, a unique gene in the rice genome, is required for tapetal development and is predominantly expressed during meiosis (Luo et al., 2006). Unigene8854 is homologous to *RTS2*, and shows peak expression at TET stage corresponding to that of *RTS2*.

Collectively, the predicted functions and verified expression patterns of the aforementioned unigenes are consistent with those of their homologues in Arabidopsis and/or in rice. This phylogenetic conservation of gene expression further validates our analysis of the *H. patens* anther transcriptome.

### 3.6 Potential genes for tapetum and pollen wall development provide useful resources for future study

Currently most characterized genes involved in tapetum or pollen wall development were identified based on genetic analysis of male sterile or reduced fertility mutants, including those with defective pollen wall development. Transcriptomic profiling is expected to expand knowledge of the genes involved in these developmental processes. In particular, our analysis of the *H. patens* anther transcriptome has allowed the identification of many more candidate genes involved in pollen development. We identified 243 differentially expressed genes and 108 stage-specific genes potentially related to tapetum layer morphogenesis and development, sporopollenin biosynthesis, exine formation, cellulose and pectin metabolism and biosynthesis, and hemicellulose and cellulose metabolism. The most significant alignments among these classes of genes derive from *Solanum lycopersicon*, *Vitis vinifera* and *Coffea canephora*. Moreover, some potential orthologues in *C. canephora* have similar annotations relevant to pollen development, for instance, callose synthase and cellulose synthase. Low-abundance transcripts expressed during anther development were also mined and differential expression patterns of genes in anthers and pollen were uncovered. This wealth of information lays the foundation for higher resolution genome-wide transcriptomic profiling of *H. patens*, functional investigation of the identified candidate genes, and the evo-devo exploration of angiosperm pollen.

#### Data Accessibility

Clean Illumina reads: NCBI SRA: SRP079620.

#### Author contributions

YFK conceptualized and designed this study. LY and YFK performed the experiments. LY, YFK, XQZ and DT performed data analysis. LY, YFK, DT and JPL wrote the manuscript. All authors have read and approved the final version of the manuscript.

#### Funding

This work was supported by National Natural Science Foundation of China (31200176, 31470410, 31670336). Guangdong Provincial Key Laboratory of Applied Botany, the National Science and Technology Infrastructure Program of China-Compilation of Ex Situ Cultivated Flora of Botanical Garden (2015FY210100), Guangdong Provincial Key Laboratory of Digital Botanical Garden and Key Laboratory of Plant Resources Conservation and Sustainable Utilization, SCBG, CAS.

#### Abbreviations:

BCP, bicellular pollen; UNM, uninucleate microspore; Ca, callose; Co, columellae; Ep, epidermis; Ec, endothecium; Ene, endexine; FL, foot layer; FT, fibrous thickenings; GN, generative nucleus; GPG, germinated pollen grain; In, intine; LD,

lipid droplets; ML, middle layer; MMC, microspore mother cell; MPG, mature pollen grain; Ms, microspores; MSA: mixed stage anther transcriptome library; N, nucleus; Nu, nucleolus; PG, pollen grains; PO, protruding onucs; PrE, primexine; PS, pericytoplasmic space; RER, rough endoplasmic reticulum; RPKM, reads per kilo bases per million mapped reads; T, tapetum; TCP, tricellular pollen; TET, tetrad; Te, tectum; Tet, tetrads; V, vacuole; VN, vegetative nucleus.

## Acknowledgments

We thank Ms. Rufang Deng and Ms. Xiaoying Hu in the SCBG for technical assistances in TEM and SEM analysis respectively, and the two reviewers for their helpful comments.

## Disclosures

The authors declare no conflicts of interest.

## References

- Aarts, M. G., Hodge, R., Kalantidis, K., Florack, D., Wilson, Z. A., Mulligan, B. J. and Pereira, A. (1997). The Arabidopsis MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. *Plant J.* 12, 615-623. doi: 10.1046/j.1365-313X.1997.00615.x
- APG III. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105-121. doi: 10.1111/j.1095-8339.2009.00996.x
- Ariizumi, T., and Toriyama, K. (2011). Genetic regulation of sporopollenin synthesis and pollen exine development. *Annu. Rev. Plant Biol.* 62, 437-460. doi: 10.1146/annurev-arplant-042809-112312
- Audic, S., and Claverie, J.M. (1997). The significance of digital gene expression profiles. *Genome Res.* 7, 986-995. doi: 10.1101/gr.7.10.986
- Aya, K., Ueguchi-Tanaka, M., Kondo, M., Hamada, K., Yano, K., Nishimura, M., et al. (2009). Gibberellin modulates anther development in rice via the transcriptional regulation of *GAMYB*. *Plant Cell* 21, 1453-1472. doi: 10.1105/tpc.108.062935
- Barrero, R.A., Chapman, B., Yang, Y., Moolhuijzen, P., Keeble-Gagnere, G., Zhang, N., et al. (2011). De novo assembly of *Euphorbia fischeriana* root transcriptome identifies prostratin pathway related genes. *BMC Genomics* 12, 600. doi: 10.1186/1471-2164-12-600
- Blackmore, S., Wortley, A.H., Skvarla, J.J., and Rowley, J.R. (2007). Pollen wall development in flowering plants. *New Phytol.* 174, 483-498. doi: 10.1111/j.1469-8137.2007.02060.x
- Blanca, J., Canizares, J., Roig, C., Ziarsolo, P., Nuez, F., and Picó, B. (2011). Transcriptome characterization and high throughput SSRs and SNPs discovery in Cucurbita pepo (Cucurbitaceae). *BMC Genomics* 12, 104. doi: 10.1186/1471-2164-12-104
- Boavida, L.C., Becker, J.D., and Feijo, J.A. (2005). The making of gametes in higher plants. *Int. J. Dev. Biol.* 49, 595-614. doi: 10.1387/ijdb.0520191b



- Bokvaj, P., Hafidh, S., and Honys, D. (2015). Transcriptome profiling of male gametophyte development. *Nicotiana tabacum. Genomic Data* 3, 106-111. doi: 10.1016/j.gdata.2014.12.002
- Borg, M. and Twell, D. (2011). Pollen: Structure and Development. In: eLS. John Wiley & Sons Ltd, Chichester. <http://www.els.net> [doi: 10.1002/9780470015902.a0002039.pub2]
- Brownfield L, Hafidh S, Borg M, Sidorova A, Mori T, et al. (2009) A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS Genet* 5(3): e1000430. doi:10.1371/journal.pgen.1000430
- Choi, H., Jin, J. Y., Choi, S., Hwang, J. U., Kim, Y. Y., Suh, M. C., et al. (2011). An ABCG/WBC - type ABC transporter is essential for transport of sporopollenin precursors for exine formation in developing pollen. *Plant J.* 65, 181-193. doi: 10.1111/j.1365-313X.2010.04412.x
- Clément, C., and Audran, J.C. (1993). Orbicule wall surface characteristics in *Lilium* (Liliaceae). An ultrastructural and cytochemical approach. *Grana* 32, 348-353. doi: 10.1080/00173139309428963
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676. doi: 10.1093/bioinformatics/bti610
- De Hoon, M.J.L., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. *Bioinformatics* 20, 1453-1454. doi: 10.1093/bioinformatics/bth078
- Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., et al. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345, 1181-1184. doi: 10.1126/science.1255274
- Dereeper, A., Bocs, S., Rouard, M., Guignon, V., Ravel, S., Tranchant-Dubreuil, C., et al. (2015). The coffee genome hub: a resource for coffee genomes. *Nucleic Acids Res.* 43, D1028-D1035. doi: 10.1093/nar/gku1108
- Dessein, S., Ochoterena, H., De Block, P., Lens, F., Robbrecht, E., Schols, P., et al. (2005). Palynological characters and their phylogenetic signal in Rubiaceae. *Bot. Rev.* 73, 354-414. doi: 10.1663/0006-8101(2005)071[0354:PCATPS]2.0.CO;2
- Deveshwar, P., Bovill, W.D., Sharma, R., Able, J.A., and Kapoor, S. (2011). Analysis of anther transcriptomes to identify genes contributing to meiosis and male gametophyte development in rice. *BMC Plant Biol.* 11, 78. doi: 10.1186/1471-2229-11-78
- Divi, U.K., Zhou, X.R., Wang, P., Butlin, J., Zhang, D.M., Liu, Q., et al. (2016). Deep sequencing of the fruit transcriptome and lipid accumulation in a non-seed tissue of Chinese tallow, a potential biofuel crop. *Plant Cell Physiol.* 57, 125-137. doi: 10.1093/pcp/pcv181
- Dukowic-Schulze, S., and Chen, C. (2014). The meiotic transcriptome architecture of plants. *Front. Plant Sci.* 5, 220. doi: 10.3389/fpls.2014.00220
- Dong, X.Y., Feng, H., Xu, M., Lee, J., Kim, Y.K., Lim, Y.P., et al. (2013). Comprehensive analysis of genic male sterility-related genes in *Brassica rapa* using a newly developed br300k oligomeric chip. *Plos One* 8, 24. doi:



10.1371/journal.pone.0072178

- Dong, X.S., Hong, Z.Y., Sivaramakrishnan, M., Mahfouz, M., and Verma, D. P. S. (2005). Callose synthase (*CalS5*) is required for exine formation during microgametogenesis and for pollen viability in *Arabidopsis*. *Plant J.* 42, 315-328. doi: 10.1111/j.1365-313X.2005.02379.x
- Dobritsa, A. A., Nishikawa, S. I., Preuss, D., Urbanczyk-Wochniak, E., Sumner, L. W., Hammond, A., et al. (2009). LAP3, a novel plant protein required for pollen development, is essential for proper exine formation. *Sex. Plant Reprod.* 22, 167-177. doi: 10.1007/s00497-009-0101-8
- Dobritsa, A. A., Lei, Z., Nishikawa, S. I., Urbanczyk-Wochniak, E., Huhman, D. V., Preuss, D., et al. (2010). LAP5 and LAP6 encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in *Arabidopsis thaliana*. *Plant Physiol.* 153, 937-955. doi: 10.1104/pp.110.157446
- Durbarry, A., Vizir, I., and Twell, D. (2005). Male germ line development in *Arabidopsis*. Duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. *Plant Physiol.* 137, 297-307. doi/10.1104/pp.104.053165
- Edlund, A.F., Swanson, R., and Preuss, D. (2004). Pollen and stigma structure and function: The role of diversity in pollination. *Plant Cell* 16, S84-S97. doi: 10.1105/tpc.015800
- El-Ghazaly, G., Huysmans, S., and Smets, E.F. (2001). Pollen development of *Rondeletia odorata* (Rubiaceae). *AM. J. Bot.* 88, 14-30.
- Feng, B.M., Lu, D.H., Ma, X., Peng, Y.B., Sun, Y.J., Ning, G., and Ma, H. (2012). Regulation of the *Arabidopsis* anther transcriptome by *DYT1* for pollen development. *Plant J.* 72, 612-624. doi: 10.1111/j.1365-313X.2012.05104.x
- Fujita, M., Horiuchi, Y., Ueda, Y., Mizuta, Y., Kubo, T., Yano, K., et al. (2010). Rice expression atlas in reproductive development. *Plant and Cell Physiol.* 51, 2060-2081. doi: 10.1093/pcp/pcq165
- Garg, R., Patel, R.K., Tyagi, A.K., and Jain, M. (2011). De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res.* 18, 53-63. doi: 10.1093/dnares/dsq028
- Goldberg, R.B., Beals, T.P., and Sanders, P.M. (1993). Anther development - basic principles and practical applications. *Plant Cell* 5, 1217-1229. doi: 10.1105/tpc.5.10.1217
- Gómez, J.F., Talle, B., and Wilson, Z.A. (2015). Anther and pollen development: A conserved developmental pathway. *J. Integr. Plant Biol.*, 57, 876-891. doi: 10.1111/jipb.12425
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644-652. doi: 10.1038/nbt.1883
- Grienenberger, E., Kim, S. S., Lallemand, B., Geoffroy, P., Heintz, D., de Azevedo Souza, C., et al. (2010). Analysis of TETRAKETIDE  $\alpha$ -PYRONE REDUCTASE function in *Arabidopsis thaliana* reveals a previously unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. *Plant Cell*, 22, 4067-4083. doi: 10.1105/tpc.110.080036

- Gu, J.N., Zhu, J., Yu, Y., Teng, X.D., Lou Y., Xu, X.F., et al. (2014). *DYT1* directly regulates the expression of *TDF1* for tapetum development and pollen wall formation in *Arabidopsis*. *Plant J.* 80, 1005-1013. doi: 10.1111/tpj.12694
- Guo, Q.Q., Ma, X.J., Wei, S.G., Qiu, D.Y., Wilson, I.W., Wu, P., et al. (2014). De novo transcriptome sequencing and digital gene expression analysis predict biosynthetic pathway of rhynchophylline and isorhynchophylline from *Uncaria rhynchophylla*, a non-model plant with potent anti-alzheimer's properties. *BMC genomics*, 15, 676. doi: 10.1186/1471-2164-15-676
- Hansson, T., and El-Ghazaly, G. (2000). Development and cytochemistry of pollen and tapetum in *Mitriostigma axillare* (Rubiaceae). *Grana* 39, 65-89. doi: 10.1080/001731300300045184
- Heslop-Harrison, J. (1968) Pollen wall development. *Science* 161, 230-237. doi: 10.1126/science.161.3838.230
- Hesse, M. (1986). Orbicules and the ektexine are homologous sporopollenin concretions in Spermatophyta. *Plant Syst. Evol.* 153, 37-48. doi: 10.1007/BF00989416
- Higginson, T., Li, S.F., and Parish, R.W. (2003). *AtMYB103* regulates tapetum and trichome development in *Arabidopsis thaliana*. *Plant J.* 35, 177-192. doi: 10.1046/j.1365-313X.2003.01791.x
- Hird, D.L., Worrall, D., Hodge, R., Smartt, S., Paul, W., and Scott, R. (1993). The anther-specific protein encoded by the *Brassica napus* and *Arabidopsis thaliana* a6-gene displays similarity to beta-1,3-glucanases. *Plant J.* 4, 1023-1033. doi: 10.1046/j.1365-313X.1993.04061023.x
- Hollender, C.A., Kang, C., Darwish, O., Geretz, A., Matthews, B.F., Slovin, J., et al. (2014). Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks. *Plant Physiol.* 165, 1062-1075. doi: 10.1104/pp.114.237529
- Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol.* 5, R85. doi: 10.1186/gb-2004-5-11-r85
- Hu, J., Wang, Z.D., Zhang, L.Y., and Sun, M.X. (2014). The *Arabidopsis* Exine Formation Defect (*EFD*) gene is required for primexine patterning and is critical for pollen fertility. *New Phytol.* 203, 140-154. doi: 10.1111/nph.12788
- Hua, W.P., Zhang, Y., Song, J., Zhao, L.J., and Wang, Z.Z. (2011). De novo transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. *Genomics* 98, 272-279. doi: 10.1016/j.ygeno.2011.03.012
- Huang, M.D., Wei, F.J., Wu, C.C., Hsing, Y.I.C., and Huang, A.H.C (2009). Analyses of advanced rice anther transcriptomes reveal global tapetum secretory functions and potential proteins for lipid exine formation. *Plant Physiol.* 149, 694-707. doi: 10.1104/pp.108.131128
- Huang, M.D., Hsing, Y.I.C., and Huang, A.H.C. (2011). Transcriptomes of the anther sporophyte: availability and uses. *Plant Cell Physiol.* 52, 1459-1466. doi: 10.1093/pcp/pcr088

- Huysmans, S., El-Ghazaly, G., and Smets, E. (1998). Orbicules in angiosperms: morphology, function, distribution, and relation with tapetum types. *Bot. Rev.* 64, 240-272. doi: 10.1007/BF02856566
- Iseli, C., Jongeneel, C.V., and Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *International Conference on Intelligent Systems for Molecular Biology* 99, 138-148.
- Jiang, J., Zhang, Z., and Cao, J. (2013). Pollen wall development: the associated enzymes and metabolic pathways. *Plant Biol.* 15, 249-263. doi: 10.1111/j.1438-8677.2012.00706.x
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480-D484. doi: 10.1093/nar/gkm882
- Kapoor, S., Kobayashi, A., and Takatsuji, H. (2002). Silencing of the tapetum-specific zinc finger gene *TAZI* causes premature degeneration of tapetum and pollen abortion in *petunia*. *Plant Cell* 14, 2353-2367. doi: 10.1105/tpc.003061
- Kuang, Y.F., Kirchoff, B.K., and Liao, J.P. (2012). Presence of the protruding oncus is affected by anther dehiscence and acetolysis technique. *Grana* 51, 253-262. doi: 10.1080/00173134.2012.660984
- Kuang, Y.F., and Liao, J.P. (2015). Calcium distribution during pollen development of *Uncaria hirsuta* Havil. (Rubiaceae), with special reference to the function of the protruding oncus. *Grana* 54, 12-25. doi: 10.1080/00173134.2014.962079
- Li, H., Yuan, Z., Vizcay-Barrena, G., Yang, C.Y., Liang, W.Q., Zong, J., et al. (2011). *PERSISTENT TAPETAL CELL1* encodes a PHD-finger protein that is required for tapetal cell death and pollen development in rice. *Plant physiol.* 156, 615-630. doi: 10.1104/pp.111.175760
- Li, N., Zhang, D.S., Liu, H.S., Yin, C.S., Li, X.X., Liang, W.Q., et al. (2006). The rice tapetum degeneration retardation gene is required for tapetum degradation and anther development. *Plant Cell* 18, 2999-3014. doi: 10.1105/tpc.106.044107
- Logacheva, M.D., Kasianov, A.S., Vinogradov, D.V., Samigullin, T.H., Gelfand, M.S., Makeev, V.J., et al. (2011). De novo sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*). *BMC genomics*, 12, 30. doi: 10.1186/1471-2164-12-30
- Loraine, A.E., McCormick, S., Estrada, A., Patel, K., and Qin, P. (2013). RNA-Seq of Arabidopsis pollen uncovers novel transcription and alternative splicing. *Plant Physiol.* 162, 1092-1109. doi: 10.1104/pp.112.211441
- Lu, L., Wortley, A.H., Li, D.Z., Wang, H., and Blackmore, S. (2015). Evolution of angiosperm pollen. 2. The basal angiosperms. *Ann. Mo. Bot. Gard.* 100, 227-269. doi: 10.3417/2012048
- Luo, H., Lee, J.Y., Hu, Q., Nelson-Vasilchik, K., Eitas, T. K., Lickwar, C., and Hodges, T. K. (2006). *RTS*, a rice anther-specific gene is required for male fertility and its promoter sequence directs tissue-specific gene expression in different plant species. *Plant Mol. Biol.* 62, 397-408. doi: 10.1007/s11103-006-9031-0
- Ma, J., Skibbe, D.S., Fernandes, J., and Walbot, V. (2008). Male reproductive

development: gene expression profiling of maize anther and pollen ontogeny. *Genome Biol.* 9, R181. doi: 10.1186/gb-2008-9-12-r181

- McCormick, S. (1993). Male gametophyte development. *Plant Cell* 5, 1265-1275. doi: 10.1105/tpc.5.10.1265
- Morant, M., Jorgensen, K., Schaller, H., Pinot, F., Lindberg Moller, B., Werck-Reichhart, D., et al. (2007). *CYP703* is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. *Plant Cell* 19, 1473-1487. doi: 10.1105/tpc.106.045948
- Ness, R.W., Siol, M., and Barrett, S.C. (2011). De novo sequence assembly and characterization of the floral transcriptome in cross- and self-fertilizing plants. *BMC Genomics* 12, 298. doi: 10.1186/1471-2164-12-298
- Owen, H.A., and Makaroff, C.A. (1995). Ultrastructure of microsporogenesis and microgametogenesis in *Arabidopsis thaliana* (L). Heynh. ecotype Wassilewskija (Brassicaceae). *Protoplasma* 185, 7-21. doi: 10.1007/BF01272749
- Pacini, E. (1997). Tapetum character states: analytical keys for tapetum types and activities. *Can. J. Bot.* 75, 1448-1459. doi: 10.1139/b97-859
- Parchman, T.L., Geist, K.S., Grahnen, J.A., Benkman, C.W., and Buerkle, C.A. (2010). Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* 11, 180. doi: 10.1186/1471-2164-11-180
- Pina, C., Pinto, F., Feijo, J.A., and Becker, J.D. (2005). Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiol.* 138, 744-756. doi: 10.1104/pp.104.057935
- Rotman, N., Durbarry, A., Wardle, A., Yang, W.C., Chaboud, A., Faure, J. et al. (2005). A novel class of MYB factors controls sperm-cell formation in plants. *Curr. Biol.* 15, 244-248. doi 10.1016/j.cub.2005.01.013
- Rutley, N., and Twell, D. (2015). A decade of pollen transcriptomics. *Plant Reprod.* 28, 73-89. doi: 10.1007/s00497-015-0261-7
- Saldanha, A.J. (2004). Java Treeview-extensible visualization of microarray data. *Bioinformatics* 20, 3246-3248. doi: 10.1093/bioinformatics/bth349
- Sanders, P.M., Bui, A.Q., Weterings, K., McIntire, K.N., Hsu, Y.C., Lee, P.Y., et al. (1999). Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. *Sex. Plant Reprod.* 11, 297-322. doi: 10.1007/s004970050158
- Scott, R.J., Spielman, M., and Dickinson, H.G. (2004). Stamen structure and function. *Plant Cell* 16, S46-S60. doi: 10.1105/tpc.017012
- Shi, C.Y., Yang, H., Wei, C.L., Yu, O., Zhang, Z.Z., Jiang, C.J., et al. (2011). Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of TET-specific compounds. *BMC Genomics* 12, 131. doi: 10.1186/1471-2164-12-131
- Shi, J.X., Cui, M.H., Yang, L., Kim, Y.J., and Zhang, D.B. (2015). Genetic and biochemical mechanisms of pollen wall development. *Trends in Plant Sci.* 20, 741-753. doi: 10.1016/j.tplants.2015.07.010

- Schnurr, J.A., Storey, K.K., Jung, H.J.G., Somers, D.A., and Gronwald, J.W. (2006). UDP-sugar pyrophosphorylase is essential for pollen development in Arabidopsis. *Planta* 224, 520-532. doi: 10.1007/s00425-006-0240-1
- Sorensen, A.M., Krober, S., Unte, U.S., Huijser, P., Dekker, K., and Saedler, H. (2003). The Arabidopsis *ABORTED MICROSPORES (AMS)* gene encodes a MYC class transcription factor. *Plant J.* 33, 413-423. doi: 10.1046/j.1365-3113X.2003.01644.x
- Souza, C.D.A., Kim, S.S., Koch, S., Kienow, L., Schneider, K., McKim, S.M., et al. (2009). A novel fatty Acyl-CoA synthetase is required for pollen development and sporopollenin biosynthesis in Arabidopsis. *Plant Cell* 21, 507-525. doi: 10.1105/tpc.108.062513
- Tsuchiya, T., Toriyama, K., Ejiri, S., and Hinata, K. (1994). Molecular characterization of rice genes specifically expressed in the anther tapetum. *Plant Mol. Biol.* 26, 1737-1746. doi: 10.1007/BF00019488
- Twell, D. (2011). Male gametogenesis and germline specification in flowering plants. *Sex. Plant Reprod.* 24, 149-160. doi: 10.1007/s00497-010-0157-5
- Verstraete, B., Moon, H.K., Smets, E., and Huysmans, S. (2014). Orbicules in flowering plants: a phylogenetic perspective on their form and function. *Bot. Rev.* 80, 107-134. doi: 10.1007/s12229-014-9135-1
- Verelst, W., Saedler, H., Münster, T. (2007a). MIKC\* MADS-protein complexes bind motifs enriched in the proximal region of late-pollen-specific Arabidopsis promoters. *Plant Physiol.* 143: 447–460. doi/10.1104/pp.106.089805
- Verelst, W., Twell, D., de Folter, S., Immink R., Saedler, H., and Münster T. (2007b). MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biology.* 8, R249. doi:10.1186/gb-2007-8-11-r249
- Vinckier, S., and Smets, E. (2002). Morphological and ultrastructural diversity of orbicules in relation to evolutionary tendencies in Apocynaceae s.l. *Ann. Bot.* 90, 647-662. doi: 10.1093/aob/mcf243
- Vinckier, S., and Smets, E. (2003). Morphological and ultrastructural diversity of orbicules in Gentianaceae. *Ann. Bot.* 92, 657-672. doi: 10.1093/aob/mcg187
- Vizcay-Barrena, G., and Wilson, Z.A. (2006). Altered tapetal PCD and pollen wall development in the Arabidopsis *ms1* mutant. *J. Exp. Bot.* 57, 2709-2717. doi: 10.1093/jxb/erl032
- Wan, L.L., Zha, W.J., Cheng, X.Y., Liu, C., Lv, L., Liu, C.X., et al. (2011). A rice beta-1,3-glucanase gene *Osg1* is required for callose degradation in pollen development. *Planta* 233, 309-323. doi: 10.1007/s00425-010-1301-z
- Wang, G., Du, X., Ji, J., Guan, C., Li, Z., and Josine, T. L. (2015). De novo characterization of the *Lycium chinense* Mill. leaf transcriptome and analysis of candidate genes involved in carotenoid biosynthesis. *Gene* 555, 458-463. doi: 10.1016/j.gene.2014.10.058
- Wang, A.M., Xia, Q., Xie, W.S., Datla, R., and Selvaraj, G. (2003). The classical Ubisch bodies carry a sporophytically produced structural protein (RAFTIN) that is essential for pollen development. *Proc. Natl. Acad. Sci. USA.* 100, 14487-14492. doi: 10.1073/pnas.2231254100



- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57-63. doi: 10.1038/nrg2484
- Wang, Z.Y., Fang, B.P., Chen, J.Y., Zhang, X.J., Luo, Z.X., Huang, L.F., et al. (2010). De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics* 11, 726. doi: 10.1186/1471-2164-11-726
- Wei, L.Q., Xu, W.Y., Deng, Z.Y., Su, Z., Xue, Y., and Wang, T. (2010). Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in *Oryza sativa*. *BMC Genomics* 11, 338. doi: 10.1186/1471-2164-11-338
- Wilson, Z.A., and Zhang, D.B. (2009). From Arabidopsis to rice: pathways in pollen development. *J. Exp. Bot.* 60, 1479-1492. doi: 10.1093/jxb/erp095
- Wortley, A.H., Wang, H., Lu, L., Li, D.Z., and Blackmore, S. (2015). Evolution of angiosperm pollen. 1. Introduction. *Ann. Mo. Bot. Gard.* 100, 177-226. doi: 10.3417/2012047
- Xu, J., Yang, C.Y., Yuan, Z., Zhang, D.S., Gondwe, M. Y., Ding, Z.W., and Wilson, Z. A. (2010). The ABORTED MICROSPORES regulatory network is required for postmeiotic male reproductive development in *Arabidopsis thaliana*. *Plant Cell*, 22, 91-107. doi: 10.1105/tpc.109.071803
- Xu, J., Ding, Z., Vizcay-Barrena, G., Shi, J.X., Liang, W.Q., Yuan, Z., et al. (2014). ABORTED MICROSPORES acts as a master regulator of pollen wall formation in Arabidopsis. *Plant Cell* 26, 1544-1556. doi: 10.1105/tpc.114.122986
- Yang, S.L., Xie, L.F., Mao, H.Z., Puah, C.S., Yang, W.C., Jiang, L., et al. (2003). The *TAPETUM DETERMINANT1* gene is required for cell specialization in the Arabidopsis anther. *Plant Cell* 15, 2792-2804. doi: 10.1105/tpc.016618
- Ye, J., Fang, L., Zheng, H.K., Zhang, Y., Chen, J., Zhang, Z.J., et al. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, W293-W297. doi: 10.1093/nar/gkl031
- Zhang, D.B., and Wilson, Z.A. (2009). Stamen specification and anther development in rice. *Chinese Sci. Bull.* 54, 2342-2353. doi: 10.1007/s11434-009-0348-3
- Zhang, D.S., Liang, W.Q., Yin, C.S., Zong, J., Gu, F.W., and Zhang, D. B. (2010). *OsC6*, encoding a lipid transfer protein, is required for postmeiotic anther development in rice. *Plant Physiol.* 154, 149-162. doi: 10.1104/pp.110.15886
- Zhang, D.B., Luo, X., and Zhu, L. (2011). Cytological analysis and genetic control of rice anther development. *J. Genet. Genomics* 38, 379-390. doi: 10.1016/j.jgg.2011.08.001
- Zhang, D.D., Liu, D., Lv, X.M., Wang, Y., Xun, Z.L., Liu, Z.X, et al. (2014). The cysteine protease CEP1, a key executor involved in tapetal programmed cell death, regulates pollen development in Arabidopsis. *Plant Cell* 26, 2939-2961. doi: 10.1105/tpc.114.127282
- Zhang, F.J., Wang, Z.Q., Dong, W., Sun, C.Q., Wang, H.B., Song, A.P., and Teng, N. (2014). Transcriptomic and proteomic analysis reveals mechanisms of embryo abortion during chrysanthemum cross breeding. *Sci. Rep.* 4, 6536. doi: 10.1038/srep06536



- Zhang, W., Sun, Y.J., Timofejeva, L., Chen, C.B., Grossniklaus, U., and Ma, H. (2006). Regulation of Arabidopsis tapetum development and function by *dysfunctional tapetum1* (*dyl1*) encoding a putative bHLH transcription factor. *Development* 133, 3085-3095. doi: 10.1242/dev.02463
- Zhang, X.M., Zhao, L., Larson-Rabin, Z., Li, D.Z., and Guo, Z.H. (2012). De novo sequencing and characterization of the floral transcriptome of *Dendrocalamus latiflorus* (Poaceae : Bambusoideae). *Plos One* 7, e42082. doi: 10.1371/journal.pone.0042082
- Zhang, Z.B., Zhu, J., Gao, J.F., Wang, C., Li, H., Li, H., et al. (2007). Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis. *Plant J.* 52, 528-538. doi: 10.1111/j.1365-3113X.2007.03254.x
- Zhao, D.Z., Wang, G.F., Speal, B., and Ma, H. (2002). The *EXCESS MICROSPOROCTES1* gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the Arabidopsis anther. *Gene. Dev.* 16, 2021-2031. doi: 10.1101/gad.997902
- Zhao, S.R., Fung-Leung, W.-P., Bittner, A., Ngo, K., and Liu, X.J. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *Plos one* 9, e78644. doi: 10.1371/journal.pone.0078644
- Zhu, E.G., You, C.J., Wang, S.S., Cui, J., Niu, B.X., Wang, Y.X., et al. (2015). The DYT1-interacting proteins bHLH010, bHLH089 and bHLH091 are redundantly required for Arabidopsis anther development and transcriptome. *Plant J.* 83, 976-990. doi: 10.1111/tpj.12942

## Tables

**Table 1 Classification of flowers and anthers for categorization of anther developmental stages.**

Pollen developmental stage	Flower length (cm)	Anther length (cm)
Microspore mother cell stage	0.30 - 0.80	0.20 - 0.45
Tetrad stage	0.90 - 1.00	0.50 - 0.60
Uninucleate microspore stage	1.10 - 1.60	0.60 - 0.85
Bicellular pollen stage	1.80 - 2.70	0.91 - 1.01

**Table 2 Annotation Results of *H. patens* Unigenes**

Sequence File	NR	NT	Swiss-Prot	KEGG	COG	GO	ALL
<i>H. patens</i> -Unigene	55,563 (61.8 %)	46,601 (51.9 %)	35,244 (39.2 %)	31,169 (34.7 %)	18,948 (21.1 %)	42,244 (47.0 %)	57,476 (64.0 %)

**Figure legends****Figure 1 Semi-thin sections of *H. patens* anthers.**

(A) Microspore mother cell stage, showing the cell layers of the anther wall and the microspore mother cells.

(B) Tetrad stage, the tapetal cells become vacuolated, the middle layer flattened.

(C) Uninucleate microspore stage, showing free microspores and cell layers of the anther wall.

(D) Bicellular pollen stage, showing pollen grains and fibrous thickenings in the endothecium.

Scale bars in A, B, C and D represent 25  $\mu$ m.

**Figure 2 Ultra-thin sections of *H. patens* anthers.**

(A) Tetrad stage, the tetrads encased by a communal callose wall.

(B) Tetrad stage, primexine matrix forms between the callose wall and the microspore.

(C) Tetrad stage, showing fragmentation of nucleoli, small vesicles and pro-Ubisch bodies (white arrowheads).

(D) Uninucleate microspore stage, showing abundant Ubisch bodies produced in the tapetum (white arrowheads).

(E) Showing a free microspore with differentiated wall.

(F) Showing white-line-centered lamella (white arrowhead) between the foot layer and the endexine.

(G) Vacuolated microspore stage, the nucleus is positioned against the cell wall.

(H) Showing the pericytoplasmic space formed beneath the exine.

(I) Showing cell layers of the anther wall and Ubisch bodies in the tapetum (white arrowheads).

Scale bars in A, C-E, G and I represent 2  $\mu$ m, B represents 100nm, F and H represent 200 nm.

**Figure 3 Ultra-thin sections of *H. patens* anthers.**

(A) Uninucleate microspore stage, showing the early-formed intine and associated rough endoplasmic reticulum.

(B) Bicellular pollen stage, showing the generative nucleus surrounded by lipid droplets.

(C) Bicellular pollen stage, showing the thickened and more electron-dense intine.  
 (D) Bicellular pollen stage, showing cell layers of the anther wall, fibrous thickenings in the endothecium and remnants of the tapetum.  
 (E) Mature pollen grains released from the anther locule.  
 (F) One pollen grain, showing the micro-reticulate ornamentation and protruding oncus..  
 Scale bars in A represents 500 nm; B represents 2  $\mu$ m; C represents 200 nm; D represents 10  $\mu$ m; E represents 50  $\mu$ m; F represents 5  $\mu$ m.

#### **Figure 4 DGE Libraries of four developmental stages.**

The white bars in this bar graph depict the numbers of genes mapped by clean reads in each library. The light grey bars depict the numbers of genes with RPKM value  $\geq 1$  in each library. The dark grey bars depict the numbers of genes with RPKM value  $\geq 5$  in each library. The black bars depict the numbers of genes with RPKM value  $\geq 10$  in each library.

#### **Figure 5 Expression patterns of highly expressed genes.**

The Venn diagram shows the constitution of genes highly expressed in at least one stage. The overlaps represent genes simultaneously highly expressed at two, three or four stages.

#### **Figure 6 Comparisons of adjacent developmental stages.**

The DGE libraries of four stages were compared to their preceding stage during anther development. MMC has no reference. The number of genes 2-fold up-regulated or down-regulated at  $FDR \leq 0.001$  are plotted on the red or black bars in the graph. Among the up-regulated genes at each stage, the numbers of specifically enriched genes (RPKM value  $\geq 10$ ) are plotted on the blue bars.

#### **Figure 7 Pairwise comparisons within four DGE libraries.**

Genes with similar expression patterns were grouped together to make 14 groups. The line graphs show the variation of expression levels at four stages of anther development.

#### **Figure 8 Stage-specifically expressed genes at four developmental stages.**

The gray bars in this bar graph depict the numbers of differentially expressed genes (genes in group 1 to 4 in Fig. 7). The black bars depict the numbers of specifically expressed genes (two-fold up-regulated and RPKM value  $\geq 10$ ) in each library.

#### **Figure 9 Expression patterns of stage-specifically expressed genes.**

Hierarchical cluster diagram represents expression patterns of stage-specifically expressed genes of four stages. The numbers of genes under each cluster are showed at the right side.

**Figure 10 Result of q-PCR analysis.**

The left Y axis represents RPKM value of each gene using RNA-Seq analysis. The right Y axis represents  $\log_2$  transformed relative transcript amount obtained by Q-PCR. The correlation co-efficient ( $r$ ) between the two expression profiles is also showed.

Fig 1

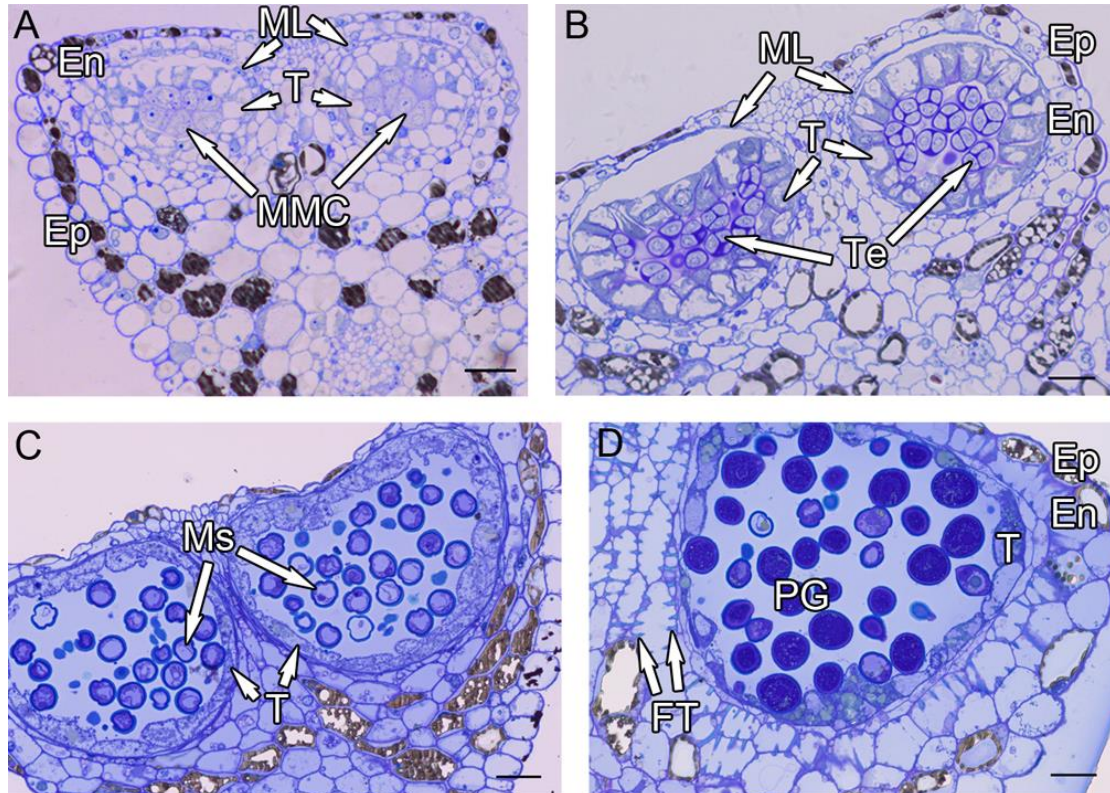




Fig 2

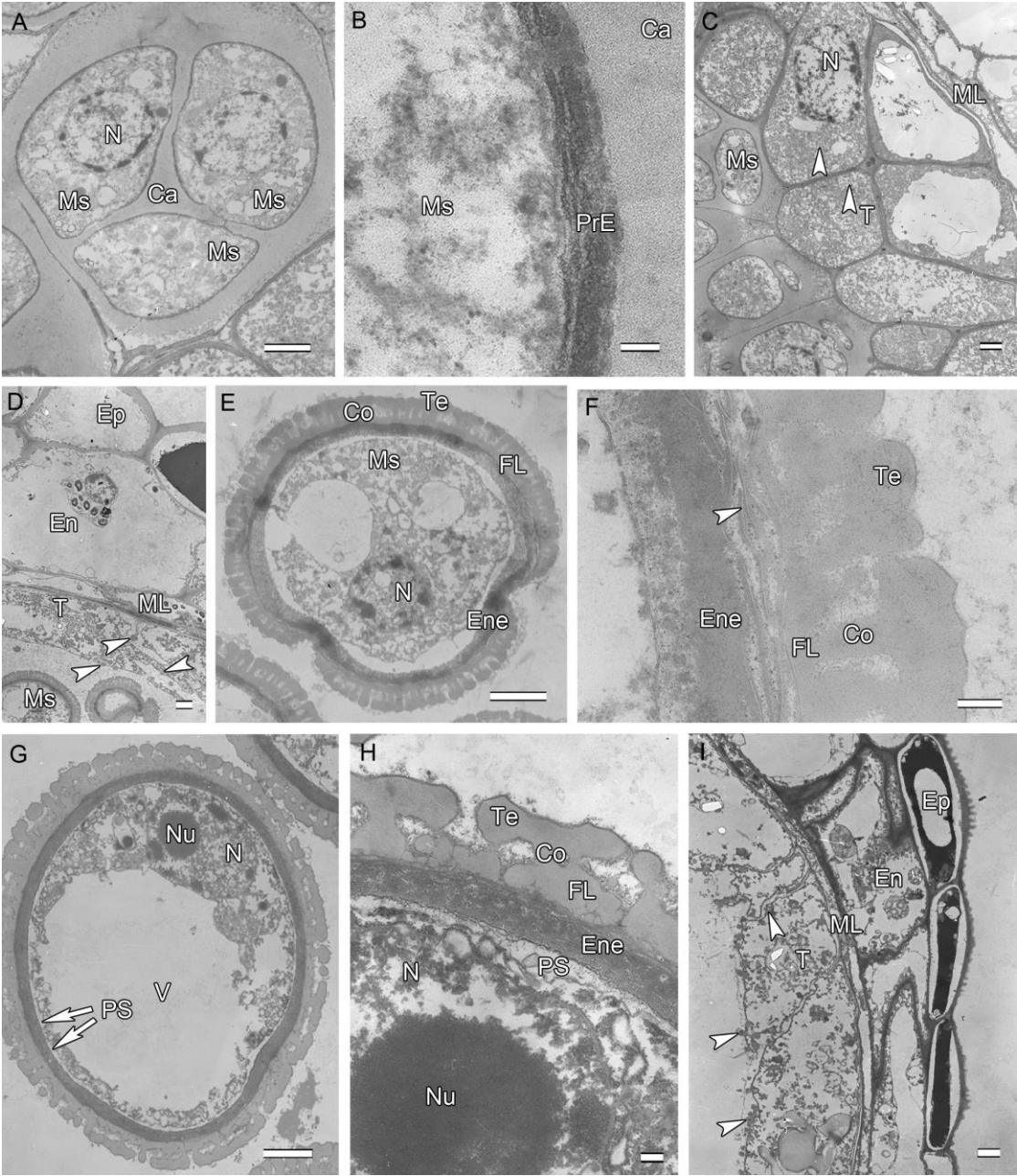


Fig 3

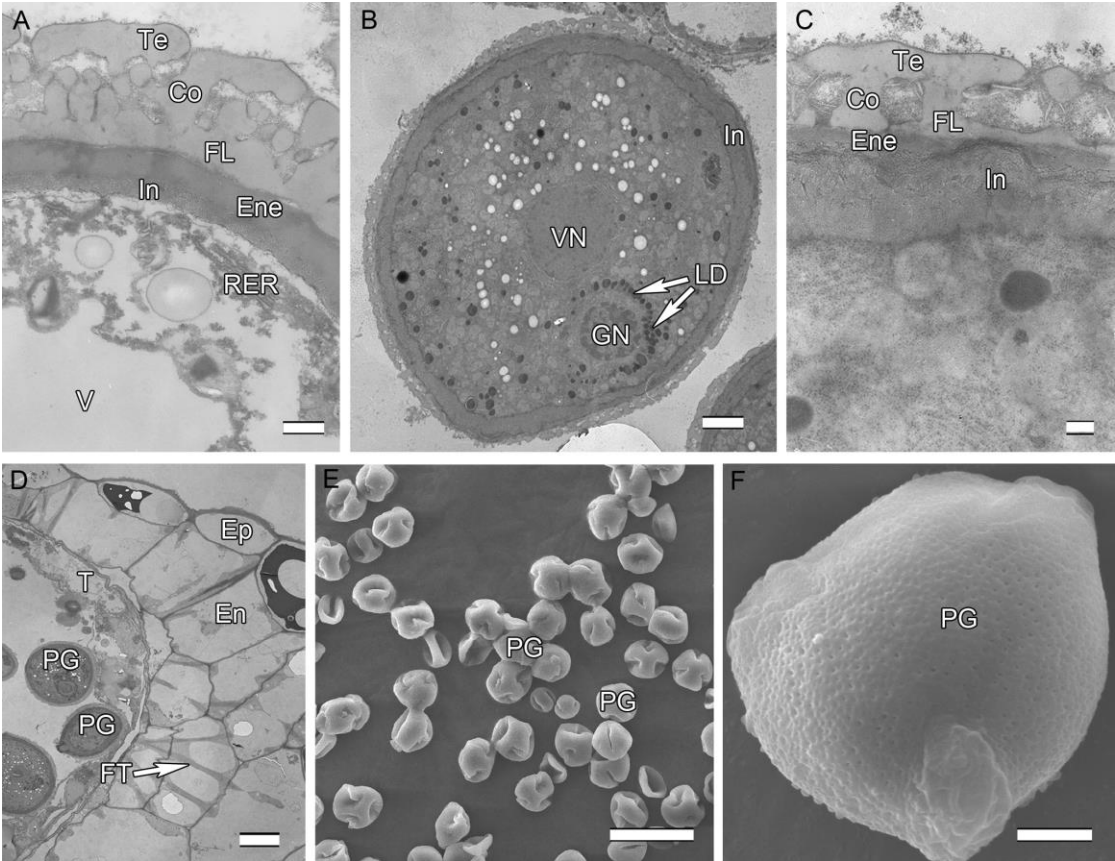


Fig 4

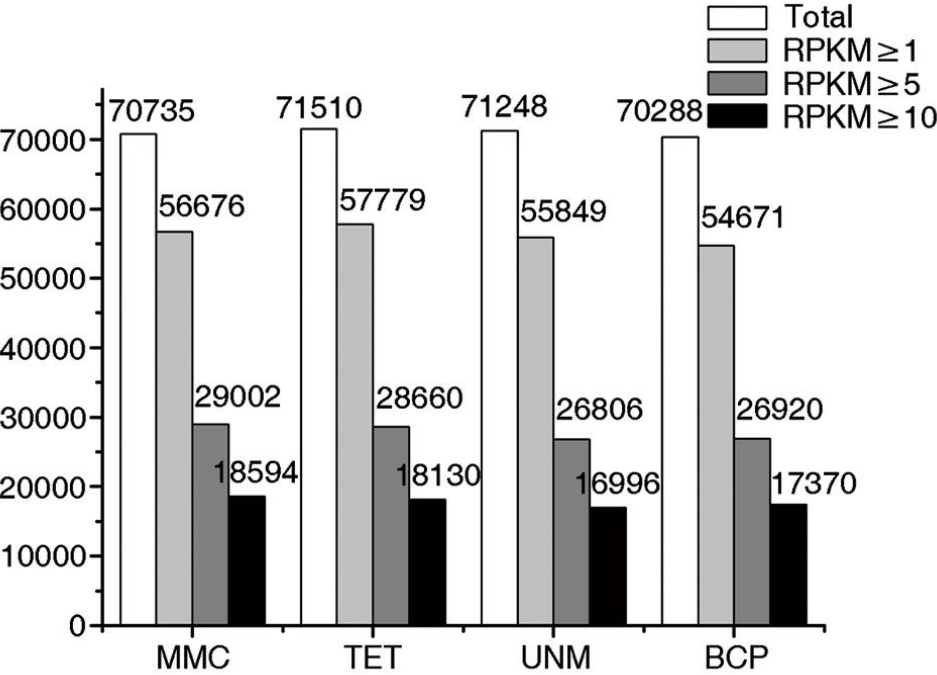




Fig 5

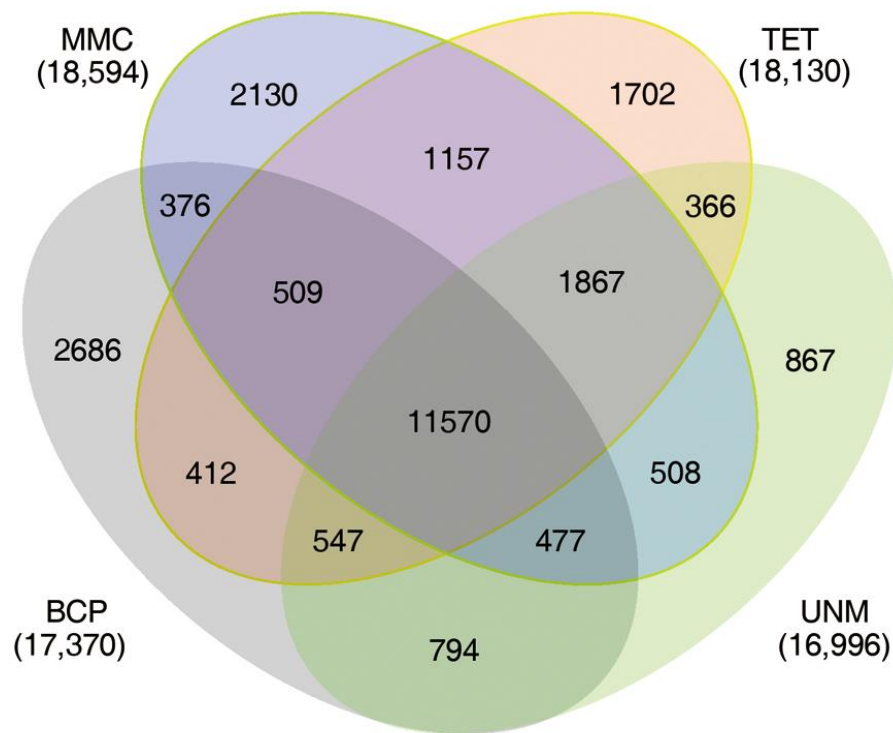


Fig 6

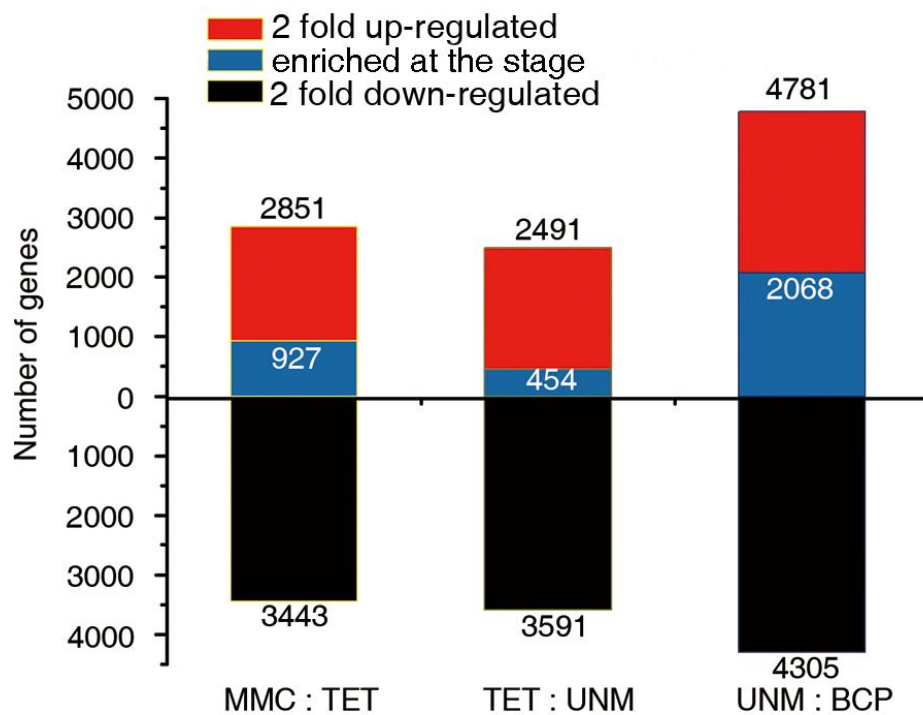


Fig 7

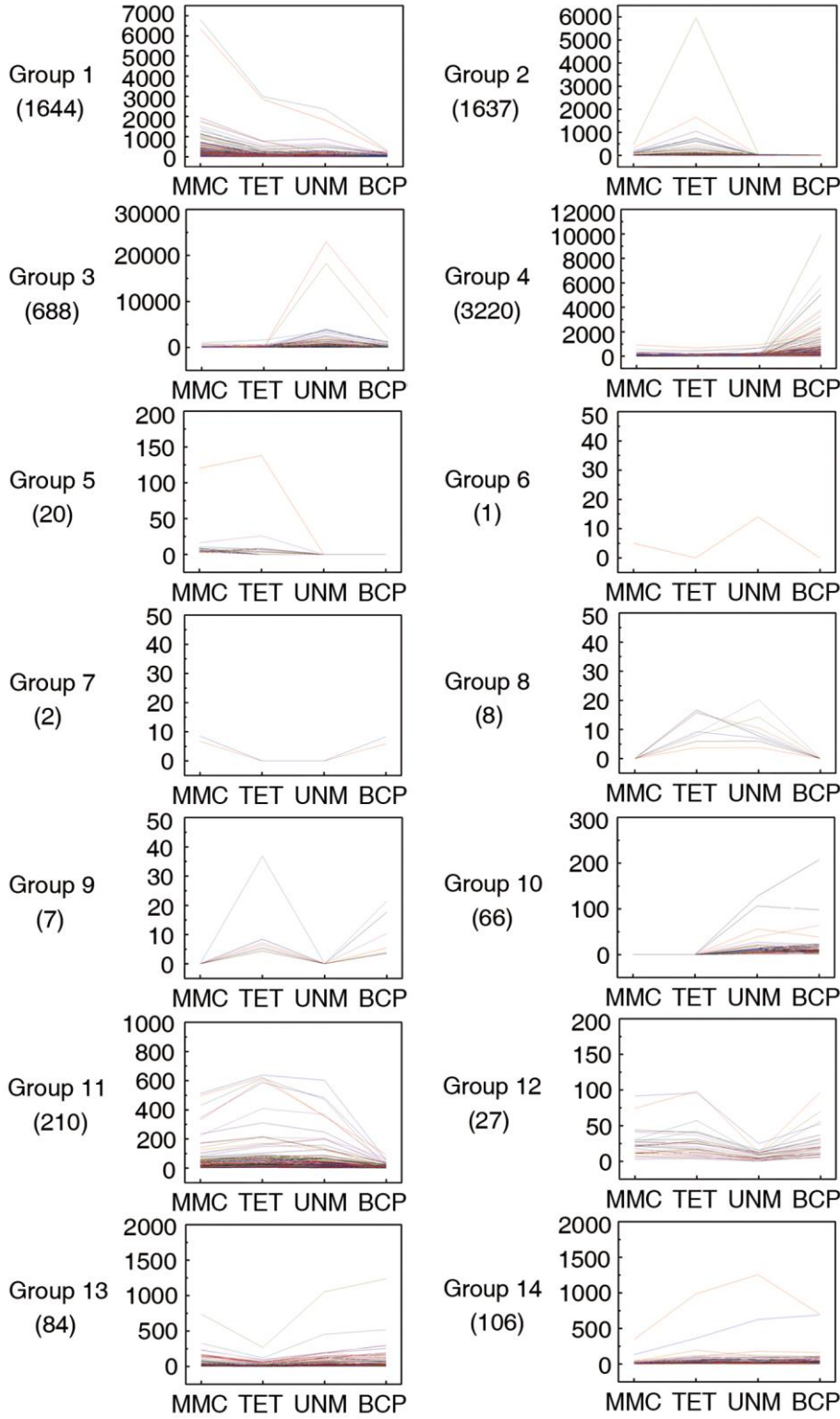


Fig 8

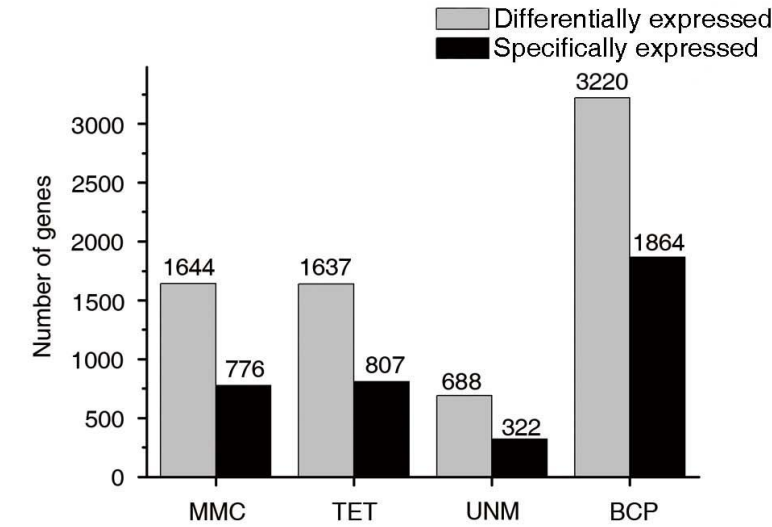


Fig 9

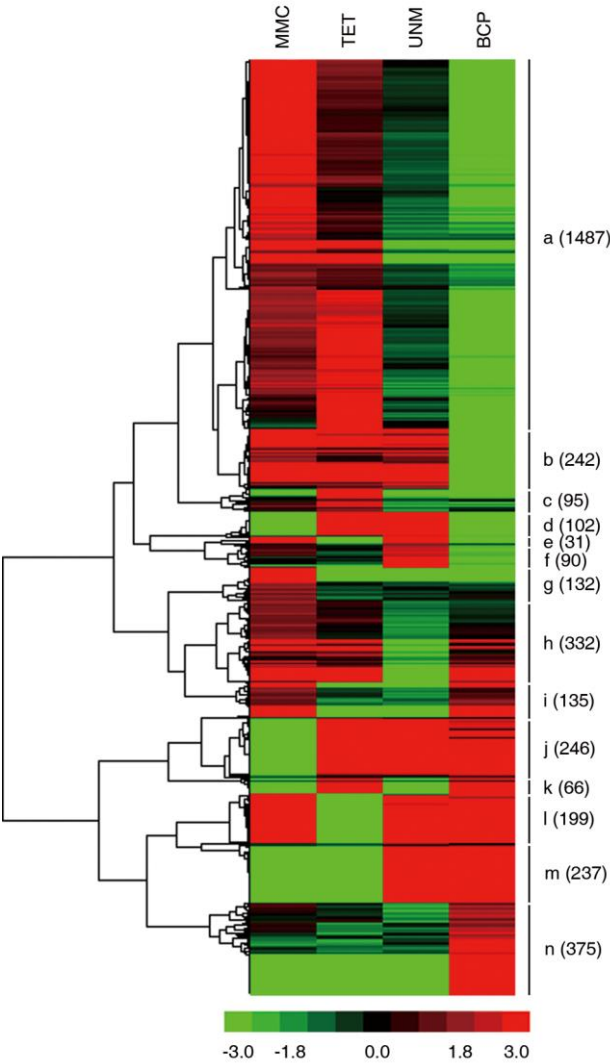


Fig 10

